UNDERGRADUATE LIBRARY

A Publication of the American Psychological Association

Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

Preference for Delayed Reinforcement and Social Responsibility 1
Walter Mischel
Stability and Fluctuation in the Power Relationships of Young Children 8
Elizabeth Gellert
The Predilection for Single Orderings 16
Clinton B. De Soto
Effects of Social Motivational Variables on Choice Reaction Time of Schizophrenics 24
Leonard D. Goodstein, Wilson H. Guertin, and Harold L. Blackburn
Word Meanings and Self-Descriptions 28
John C. Loehlin
The Effectiveness of Unanticipated Persuasive Communications 35
Jane Allyn and Leon Festinger
Personality and Verbal Conditioning Effects 41
Georgia Babladelis
Attitudes toward Old People: The Development of a Scale and an Examination of Correlates 44
Nathan Kogan
The Relationship between Medical and Psychiatric Symptoms 55
Ruth G. Matarazzo, Joseph D. Matarazzo, and George Saslow
Analysis of the Connotative Meanings of a Variety of Human Values as Expressed by American College Students 62
Charles E. Osgood, Edward E. Ware, and Charles Morris
Reduction of Cognitive Dissonance by Seeking Consonant Information 74
J. Stacy Adams

Continued on inside front cover

Volume 62 JANUARY 1961 Number 1

CONTENTS—*Continued*

Familial Correlates of Aggression in Nondelinquent Male Children 79
William McCord, Joan McCord, and Alan Howard

Self-Concept Differences in Relation to Identification, Religion, and Social Class 94
James Bieri and Robin Lobeck

Emotional Factors in the Forgetting of Word Associations 99
George Levinger and James Clark

Responsiveness of Chronic Schizophrenics and Normals to Social and Nonsocial Feedback 106
Walter J. Johannsen

Interpersonal Decision Making: The Resolution of a Dyadic Conflict 114
John M. Atthowe, Jr.

Cognitive Controls of Attention Deployment as Determinants of Visual Illusions 120
Riley W. Gardner

Modification of Behavior through Verbal Conditioning: Effects in Group Therapy 128
Leonard P. Ullmann, Leonard Krasner, and Beverly J. Collins

The Effects of a Communication Attitude Change and Scale Judgments 133
Walter Weiss

CRITIQUE AND NOTES

Time Orientation and Social Class: A Correction 141
Joel E. Greene and Alan H. Roberts

Need Achievement, Fear of Failure, and Selective Recall 142
Walter R. Reitman

Value, Magnitude, and Accentuation 145
Samuel E. Stayton and Morton Wiener

Dogmatism and Learning 148
Howard J. Ehrlich

Some Contrast Effects in Social Perception 150
Douglas S. Holmes and Leonard Berkowitz

A Note on Anxiety, Instructions, and Word Association Performance 153
Irwin G. Sarason

Some Verbal Aspects of Primary Process Thought in Schizophrenia 155
Alvin G. Burstein

Some Inconsistencies in the Effect of Motivation Arousal on Humor Preferences 158
Donn Byrne

The Effect of Induced Stress on Free Associations 161
Eli Saltz

The Effects of Anxiety and Threat on the Solution of a Difficult Task 165
Irwin G. Sarason

Ego Strength and Perceived Competence as Conformity Variables 169
Kay H. Smith

The "Yeasayer" and Social Desirability: A Comment on the Couch and Keniston Paper 172
James B. Taylor

A Note on the Couch and Keniston Measure of Agreement Response Set 173
Allen L. Edwards and Jerald N. Walker

Agreeing Response Set and Social Desirability 175
Arthur Couch and Kenneth Keniston

Social Desirability and Agreement Response Set 180
Allen L. Edwards and Jerald N. Walker

A Note on Rokeach's Theory of Prejudice 184
Harry C. Triandis

Belief versus Race as Determinants of Social Distance: Comment on Triandis' Paper 187
Milton Rokeach

Volume 62 January 1961 Number 1

Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *University of California, Berkeley*
Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

ROGER BROWN
Massachusetts Institute of Technology

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
System Development Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

DONALD W. FISKE
University of Chicago

N. L. GAGE
University of Illinois

ROGER W. HEYNS
University of Michigan

ROBERT R. HOLT
New York University

IRVING L. JANIS
Yale University

HAROLD H. KELLEY
University of Minnesota

LEONARD S. KOGAN
Community Service Society of New York

RICHARD S. LAZARUS
University of California, Berkeley

DANIEL J. LEVINSON
Harvard Medical School

DAVID C. MCCLELLAND
Harvard University

LEO POSTMAN
University of California, Berkeley

ELIOT H. RODNICK
University of California, Los Angeles

NEVITT SANFORD
University of California, Berkeley

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
University of Pennsylvania

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The subscription price per volume is $10.00, or $20.00 annually (foreign, $20.50). Single issues are $4.00

Original manuscripts should be sent to the Editor-elect:

DANIEL KATZ
Center for Advanced Study in
the Behavioral Sciences
202 Junipero Serra Boulevard
Stanford, California

(After August 15, 1961)
Department of Psychology
University of Michigan
Ann Arbor, Michigan

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-class postage paid at Baltimore, Maryland, and at additional mailing offices

Copyright 1961 by the American Psychological Association, Inc.

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 1-7

PREFERENCE FOR DELAYED REINFORCEMENT AND SOCIAL RESPONSIBILITY[1]

WALTER MISCHEL[2]
Harvard University

In an earlier study (Mischel, 1958) the writer applied to cross-cultural research a technique for eliciting preference for immediate, smaller as opposed to delayed, larger reinforcements (in the form of candy varying in size, with the smaller obtainable immediately and the latter requiring a week's waiting period). Trinidadian children aged 7 to 9 were the subjects. The main finding was a significant relationship between father absence and preference for immediate reinforcement. The present study is designed to explore some further antecedents and correlates of such choice preferences. Again the subjects were drawn from the Trinidadian culture, and all testing was conducted within the field situation.

Numerous nonexperimental references have been made to the importance of tolerance or preference for delayed rewards, and the necessity for reward delay in a multitude of complex human situations has long been recognized: it is reflected in such theoretical formulations as the "pleasure principle" and the "reality principle" (Freud, 1922). "Psychopathy," immaturity, and criminal and neurotic behavior have been seen as at least partially explicable in terms of the inability to postpone immediate gratification for the sake of delayed rewards (e.g., Mowrer & Ullman, 1945). In spite of this, there exists a relative paucity of directly relevant experimentation. Recently, a number of investigators have turned to the study of "delay capacity" (Singer, 1955), primarily as inferred from the Human Movement response (*M*) on the Rorschach, and its behavioral correlates (e.g., Levine, Glass, & Meltzoff, 1957; Levine, Spivack, & Wight, 1959; Meltzoff & Levine, 1954; Singer, Wilensky, & McCraven, 1956; Spivack, Levine, & Sprigle, 1959), such as planfulness and intelligence. In contrast to these studies, the present research approaches the problem of delay of gratification by using direct choice measures of preference for immediate, smaller or delayed, larger rewards in particular choice situations and seeks to explore the antecedents and correlates of such choices.

The major aim of the present report is to explore the relationship between preference for immediate, smaller as opposed to delayed, larger rewards and social responsibility. It was hypothesized that subjects preferring immediate, smaller rewards would show less social responsibility than subjects preferring larger, delayed rewards. Social responsibility was determined by responses to a quite carefully constructed and independently validated Social Responsibility Scale (Harris, 1957; Harris, Clark, Rose, & Valasek, 1954) as well as by the behavioral criterion of committed delinquency. The Social Responsibility Scale (SRS) used has been found to correlate substantially with other measures of personal and social adjustment and was designed to "discriminate children who have, with their peers, a reputation for responsibility as contrasted with children who have little reputation for responsibility" (Harris, 1957, p. 326). Harris' theory of responsibility, to which the items in the scale were related, conceptualized responsibility "as a composite of attitude elements reflecting behavior classifiable as reliable, accountable, loyal, or doing an effective job" (Harris, 1957, p. 322).

In the present study it was predicted that the SRS would discriminate subjects varying in preference for immediate, smaller or delayed, larger rewards, with those choosing the delayed, larger rewards showing greater social responsibility. The underlying reasoning is as follows. Within the general framework of Rotter's social learning theory (Rotter, 1954) the subject's choice behavior may be taken as a function of his expectation that the reward will actually occur (i.e., that the particular promise

[1] This study was carried out under the support of a grant from the Laboratory of Social Relations, Harvard University.

[2] Grateful acknowledgement is also due to the Education Office of the Government of Trinidad and Tobago, and to the cooperation of its officials.

will be kept) and the reward value, or preference value, of the particular reinforcement. Previous research (Mahrer, 1956) has already shown that the subject's expectancy that even after time delay reinforcement would follow from the social agent making the promise is an important variable in this kind of choice behavior. The orientation of the present research similarly emphasizes the subject's expectancy for reinforcement to follow as a function of making a particular choice with a particular social agent in a given situation, rather than the time interval itself. In common sense terms, this kind of choice behavior may be thought of as "trust" or the belief that the agent promising delayed reinforcement will actually supply it.

Such trust or willingness to postpone immediate smaller gratification for the sake of a later but larger reward in a choice situation may be regarded as a basic ingredient of social responsibility and as such should be (*a*) positively related to social responsibility scores as elicited by the SRS, and (*b*) negatively related to delinquent or criminal behavior. The SRS in turn, having been correlated with a series of measures of adjustment and maturity, may well supply a gross index of a set of variables that could not be individually manipulated within the vicissitudes of the present field research situation, but the overall relationship of which to preference for delayed reward requires empirical examination. A third, and much more tentative and minor prediction was made, anticipating a positive relationship between accuracy in the recall of the time of a past event with preference for delayed reward. The measure of time recall accuracy was included with a part of the sample as a step in an exploratory attempt to ascertain relations between preference for delayed reward and time judgments, it being anticipated that subjects preferring the larger, delayed rewards would show greater accuracy. The rationale for this is too elaborate and tentative for presentation here (Epley & Ricks, 1959). However, since recent research (Spivack, Levine, & Sprigle, 1959) has demonstrated relations between "delay capacity" (measured primarily by the Rorschach *M* response) and time estimation scores, the presently available data warrant reporting. The two major hypotheses are directed a testing the linkage between delay of reinforcement and the concepts of social irresponsibility, "psychopathy," and criminal behavior whose inverse relationships to the ability t delay immediate rewards for the sake of later larger gains have often been assumed without adequate demonstration.

METHOD

Subjects. A total of 206 subjects, all Trinidadia Negro children aged 12, 13, and 14, with a modal ag of 13, were tested. Of these, 136 (78 boys, 58 girls came from a large elementary government school (GS on the outskirts of the capital city, and 70 boys cam from a "boys' industrial school" which is a refor school for committed juvenile delinquents (JD), als located on the outskirts of the capital. Accurate inte pretation requires a more detailed description of th two schools from which subjects were drawn an within which they were tested. GS is a governme "elementary school" (primary grades with the additio of the American equivalent of the junior high scho grades), containing the most "representative" sampl of relatively urbanized, Negro, lower class children, o the basis of advice from the local school authoritie and several inspection trips to schools in the area. Th latter tended to confirm the impression given by th authorities. Approximately one third of the childre at GS come from ethnic backgrounds other tha Trinidadian Negro (e.g., East Indian, Chinese); a similar proportion with respect to different ethn backgrounds holds at JD, and, indeed, applies at th national level. All subjects selected for the experime proper at both schools were Trinidadian Negroes, ha no further restrictions, other than age (all subjec ranging from 12 to 14) and availability at the time o testing were imposed in the selection of subject Examination of the family backgrounds, e.g., famil size and constellation, living conditions, parent occupations and income, etc., of 10 randomly selecte children from each of the two schools suggested maj similarities with respect to the gross external aspec of living and economic circumstances in the bac grounds of the GS and JD groups. JD contains approx mately 250 boys, ranging in age from 10 to 18, wh have been committed for extremely heterogeneo reasons and offenses, with a minimum of two previo convictions for delinquency prior to commitment, th most common offense being petty larceny. The physic facilities at the school are relatively good and th *physical* conditions (with respect to food, clothing, an living quarters) compare favorably on the whole wi those of noninstitutionalized children from the low middle and lower socioeconomic strata in Trinida (e.g., GS). Specific inquiry was made into the ava ability of candy for the children at JD, since this relevant to the experimental procedure. This inquir

[3] There were no differences approaching statistic significance in the proportion of subjects obtained fro each age level at JD as compared to GS.

revealed that the children are allowed to have and eat candy freely, that they have "allowances" sufficient for them to purchase a large variety of candy on their own three or four times weekly, and that this is done regularly by virtually all of the children. This situation is again highly comparable to that of the sample from GS with respect to the subjects' familiarity with, and use of, candy. The school is residential, with the children living in dormitories, 40 to 50 per dormitory, assigned on an age basis.

Measures. An actual behavior choice that elicited preference for either immediate, smaller reinforcement or delayed, larger reinforcement was used at each school (see below). In addition, two direct questions were added at JD in the attempt to obtain a preliminary test of the consistency with which subjects prefer immediate or delayed rewards, and in the attempt to relate the consistency of such preferences to the dependent measures. The two questions were inserted as Items 5 and 15 of the shortened SRS form administered at JD and were phrased as follows:

1. I would rather get ten dollars right now than have to wait a whole month and get thirty dollars then.
2. I would rather wait to get a much larger gift much later rather than get a smaller one now.

The behavioral choice of preference for immediate, smaller as opposed to delayed, larger reinforcement was given as the last item of the total battery.

All items from the 50-item SRS (Harris, 1957, called "Social Attitudes Scale," p. 324, Table 1), with the exception of Item 50, were administered at GS. This item was omitted on the basis of anticipated inappropriateness or irrelevance due to cultural differences, both as directly observed in the course of field work and on the basis of advice from local informants familiar with the culture. Minor modifications in phrasing were employed on six other items to make them culturally appropriate.[4] The size of the scale was drastically reduced for administration at JD since the administration at GS (which was conducted first) revealed that the procedure was too long and restlessness on the part of the subjects was imminent. It was anticipated that subjects at JD would show even less attention for any prolonged time period and the number of items was consequently further reduced to shorten the procedure. An essentially random procedure was used for dropping items, retaining, however, a similar proportion of items with personal vs. nonpersonal reference. The SRS constituted 18 items of the 25-item revised battery: Items 1, 2, 3, 7, 11, 17, 19, 23, 24, 27, 28, 29, 31, 32, 40, 43, 44, and 48 of the SRS as reported by Harris (1957).

Subjects at JD were also asked, as one of the items in the series of questions, to record the year during which the national election took place. This event, occurring in 1956, was of major importance to most Trinidadians since it marked the beginnings of a new government, the first election of its kind, the first step towards Federation, etc., and continued to be a matter of major popular concern at the time of the study (1959). All the children had heard of the election and were, in general terms, familiar with its occurrence: the task here was to place it correctly in time. This served as the measure of accuracy in the recall of a past event.

The reinforcements. Following the procedure developed in an earlier study (Mischel, 1958) pre-experimental sessions were used to select two appropriate reinforcements for use in the experiment proper. Towards this end, 20 Negro boys randomly selected from another group of Trinidadian delinquent subjects living within the capital city and of comparable ages, were seen in individual sessions and their preferences for various specific reinforcements were elicited. As a result, two reinforcements (both candy, but varying markedly in price, size, and packaging, i.e., a $.05 and a $.25 candy bar) were selected. These met the desired requirements inasmuch as the larger reinforcement was uniformly preferred in a straight choice situation ("Which *one* of these two would you like to take?"), but when the choice was "You can have this one (the smaller) today *or* this one (the larger) in one week," approximately 50% of the group chose the former and approximately 50% the latter. The purpose of this procedure was to select a reinforcement pair that would, as closely as possible, dichotomize the choices of subjects in the JD group, without actually testing within the JD setting to avoid the anticipated contamination effects arising from pre-experimental discussions by the subjects in the same residential school.

Procedure. The experimenter was introduced as an American from a college in the United States who was interested in gathering information on the children in the various schools of the island.[5] To help with this the subjects were asked to answer a number of questions. The following statement was included in the general introduction:

> Let me explain to you what we are going to do today. It is really quite simple if you will listen very carefully. As a part of our studies we are collecting information of all sorts from children and students all over the world. None of the things you will be doing are tests or examinations . . . none of them have right or wrong answers . . . none of them will be marked or graded . . . and none of them will in

[4] The scale used was identical with that presented in Table 1, p. 324 of the Harris (1957) study, with Item 50 omitted and the following items modified to read: 18. When I work *in a group* (instead of "on a committee") I usually let other people do most of the planning; 32. Never give a *fool* (instead of "sucker") an even chance; 35. It's a good thing the *Ocean* (instead of "Atlantic ocean") separates us from Europe because then we don't have to worry about them; 40. If everyone *helps* (instead of "pitches in") to do a job, it can always get done; 42. Doing things which are important should come before things you enjoy doing *or like to do*; 49. When given a *job* (instead of a "task") I stick to it even if things I like to do better come along. The italicized words above indicate the new modifications made in the items.

[5] The same experimenter was used in all testing and every effort was made to create similar testing atmospheres in all sessions.

any way affect your own work here at school. As a matter of fact, your teachers here will never see your answers to any of the questions that we will be asking you. Nothing will be marked or graded. Instead, we want to get some information about you . . . like how old you are and how you feel about different things. Although there are no right or wrong answers, it is *most important* that you try to be honest and truthful throughout: we are interested in how and what *you* are feeling and thinking . . . . Remember, this is *not* an examination, there are no right or wrong answers, your teachers will not see what you write . . . and you will help us most by being completely truthful and honest throughout.

All questions were deliberately posed orally, with the subjects following the experimenter question by question on the printed questionnaire, since previous work in the same culture strongly indicated the superiority of this method with respect to insuring greater attention, interest, and clarity. Subjects were carefully instructed about how to record their answers next to the appropriate numbers on their questionnaires. The experimenter then said:

Now I will read each of the sentences, one at a time, and repeat each sentence once to make sure you understand . . . . After I have read the sentence two times I want you to put a Y for *Yes, I agree with it* down if you *really* do agree with the sentence, that is, if it is true for you. But put down an N for *No* if you do not agree with it or if it is not true for you. For example, if I read: I am a boy who lives in Trinidad, all the boys would put down a Y and all the girls would put down an N. Do you understand? Make sure that you keep up with me on your paper and that you put your first Y or N down next to the Number 1, your second next to the Number 2, and so forth. Now, we will read one sentence at a time, I will repeat each sentence once, and you will put down a Y for Yes if you agree with it, it is true for you, and an N for No if you disagree with it or it is not true for you. Some of the sentences will be true for you and some will not. Now let us read the first one, listen carefully.

Upon completion of the questions, the experimenter expressed his wish to thank the group for their cooperation. He displayed the two kinds of reinforcement and said:

I would like to give each of you a piece of candy but I don't have enough of these [indicating the larger, more preferred reinforcement] with me today. So you can either get this one [indicating the smaller, less preferred reinforcement] right now, today, or, if you want to, you can wait for this one [indicating] which I will bring back next Friday [one week delay interval].

These instructions were repeated in rephrased form to insure clarity and both reinforcements were carefully displayed. It was stressed that getting the (smaller) candy today precluded getting the (larger) one next week and vice versa. Subjects were asked to indicate their choice by writing T (today) or F (Friday, next week) on their questionnaires. The response made here was the measure of the behavioral choice of a larger (or more preferred) delayed reinforcement or a smaller (less preferred) immediate reinforcement. Subjects were seated sufficiently far apart from each other to insure reasonably that their choices were made independently. All testing was done in a group setting and conducted on the same weekday (but in different weeks) and at similar times of day, within each of the schools, in large classrooms.

RESULTS AND DISCUSSION

Examination of the relationship between sex and preference for delayed or immediate reinforcement revealed no differences approaching statistical significance and consequently the male and female data were treated collectively in subsequent analyses. Comparison of the delinquent (JD) and nondelinquent (GS) groups on the experimental choice was based on the following data. At JD, 31 subjects chose the immediate reinforcement and 39 chose the delayed reinforcement; at GS, 35 subjects chose the immediate and 101 subjects chose the delayed. Application of the chi square test to these data yielded a chi square of 6.48 ($p < .02$), a significantly larger proportion of delinquent subjects choosing immediate, smaller reinforcement, as was predicted.[6]

Turning to the within-group data, the following results were obtained. At GS, comparison of SRS scores of subjects choosing immediate, smaller reinforcement with those choosing delayed, larger reinforcement (Table 1) resulted in a t of 2.24 ($p < .025$), with differences in the predicted direction, i.e., subjects preferring the delayed, larger reinforcement have significantly higher social responsibility scores.

At JD, where three measures of reinforcement preference were used, the consistency of such preferences (Table 2) can be examined. The responses to the two questionnaire items eliciting preference for immediate, smaller or delayed, larger reinforcement were grouped into the following categories: consistent immediate reward (both questions answered in the direction of preference for immediate, smaller reinforcement); consistent delayed reward (both questions answered in the direction of preference for delayed, larger reward); and inconsistent responses (one

[6] All chi square tests on 2 × 2 contingency tables are corrected with Yates correction for continuity throughout this study.

TABLE 1

RELATIONSHIP BETWEEN REINFORCEMENT PREFERENCE AND SOCIAL RESPONSIBILITY SCALE (SRS)

Group	Reinforcement Preference	SRS					
		N	*M*	*SD*	Item mean[a]	*t*	*p*[c]
JD (delinquents)	Immediate (candy choice)	31	8.1	2.51	.45		
	Delayed (candy choice)	39	9.0	2.26	.50	1.55	< .10
	Consistently immediate[b]	12	7.17	2.61	.40		
	Consistently delayed[b]	19	9.11	2.32	.51	2.04	.025
GS (nondelinquents)	Immediate (candy choice)	35	28.77	3.29	.59		
	Delayed (candy choice)	101	30.43	3.92	.62	2.24	< .025

[a] SRS mean divided by the number of items in the scale: 18 at JD, 49 at GS.
[b] Over the three measures.
[c] One-tailed test of significance.

TABLE 2

CONSISTENCY OF PREFERENCE FOR IMMEDIATE OR DELAYED REINFORCEMENT: COMPARISON OF ACTUAL CHOICE WITH THE TWO QUESTIONNAIRE RESPONSES AT JD

Actual Choice[a]	*N* Consistently Immediate[b]	*N* Consistently Delayed[c]	*N* Inconsistent
Immediate	12	4	15
Delayed	4	19	16[d]

[a] Candy situation.
[b] Both questionnaire items answered in the direction of preference for immediate reward.
[c] Both questionnaire items answered in the direction of preference for delayed reward.
[d] $\chi^2 = 13.07$; $p < .01$.

question answered in the direction of immediate reward and the other in the direction of delayed reward). The responses falling within each of these categories were in turn compared with the subject's actual choice in the behavioral preference situation. Applying the chi square test to this table yielded a chi square of 13.07 ($.01 > p > .001$); there is, thus, a more than random association between responses made to each of the three measures used. Since the three measures have much in common in their phrasing and were all obtained within the same testing situation, these data provide only a limited, first step in the appraisal of the intra-individual consistency of such reinforcement preferences.

At JD, a *t* test comparing the SRS scores of subjects choosing immediate, smaller reinforcement in the actual behavior choice (candy situation) with those choosing delayed, larger reinforcement (Table 1) indicates that the difference in the mean scores of the two groups, although in the predicted direction, does not reach statistical significance. For this group, however, it is possible to examine the relationship between relatively consistent reinforcement preferences and SRS scores. Table 1 compares the mean of the SRS scores of subjects showing a consistent preference for immediate, smaller reinforcement with the mean of those consistently preferring larger, delayed reinforcement and this comparison yields a *t* (one-tailed test) of 2.04, with the differences between means being significant ($p < .025$) in the anticipated direction. Thus, although the candy choice behavior in itself does not differentiate subjects within this group with respect to SRS, the classification of subjects into relatively more consistent reinforcement preference patterns through the use of multiple preference measures does yield the predicted differences. This finding strongly suggests that future research exploring the antecedents and correlates of preference for immediate or delayed reward might more fruitfully employ multiple rather than single measures of reinforcement preference.

Table 3 shows the relationship between accuracy in the recall of the time of a past event and reinforcement preferences at JD. Responses to the question eliciting the year of the last national election were categorized into "accurate" (1956) and "inaccurate" (all other responses). Applying the chi square test to the data in Table 3 results in a significant relationship between the choice behavior and accuracy of time recall. Subjects preferring the immediate smaller reward in the candy choice situation show less accuracy in time recall ($\chi^2 = 6.35$, $p < .02$).

TABLE 3
REINFORCEMENT PREFERENCE AND ACCURACY OF TIME RECALL AT JD

Reinforcement Preference	Accuracy of Time Recall			
	N Accurate	*N* Inaccurate	χ^2	*p*
Immediate choice[a]	4	27		
Delayed choice[a]	17	22	6.35	<.02
Consistently immediate[b]	4	8		
Consistently delayed[b]	15	4	4.86	<.05

[a] Choice in candy situation only.
[b] Consistent preference over the three measures.

This finding is relevant to the work of Spivack, Levine, and Sprigle (1959) who related the length of time estimates made by subjects asked to state when 15″, 30″, and 60″ time intervals had elapsed to other measures of "ego delay function" and to intelligence. The possible theoretical significance of such relationships between particular patterns of preference for immediate or delayed rewards and time judgments, time span (Epley & Ricks, 1959), and other measures of the subjects' conceptions and uses of time, cannot be elaborated in the present report but is the object of ongoing research. At present, this finding provides only another provocative empirical correlate of the choice behavior of interest. It should be noted, however, that a recent study (Melikian, 1959) has indicated relations between a preference for delayed reinforcement measure similar to that used in the present study and intelligence, and that scores on our time recall measure may be more dependent upon the subject's intelligence than upon his accuracy specifically in making time statements. In Table 3, comparison of the accuracy of time recall scores of the subjects consistently preferring immediate or delayed reward with the total figures based on the candy choice only suggests that the major difference is that those making consistent reward choices over the three measures, whether for delayed or immediate reward, give the preponderance of "right" responses to the time recall question for the total sample. As such, the time recall measure may indeed reflect the subject's alertness, memory, consistency, and other such intellectual functions which in turn may be related to response consistency over the three measures.

Finally, the data permit some comparison of SRS scores of Trinidadian subjects with the scores obtained in the United States standardization procedures on overlapping age groups. The differences in mean SRS scores obtained in the GS (nondelinquent) sample, as compared to those reported by Harris in 1957 for the United States groups (for boys, *M* of SRS at age 12 = 36.94, *SD* = 5.14; *M* of SRS for girls at age 12 = 39.42, *SD* = 3.99) are so great as to make tests for the significance of the difference superfluous. Clearly (and recognizing that the present data are based on a 49-item scale whereas Harris' are based on 50 items) the Trinidadian subjects showed significantly lower social responsibility scores on the modified United States scale. This finding is fully consistent with anthropological observations made about the culture on the basis of three field trips and well supported by numerous other workers in the Caribbean (e.g., Herskovits & Herskovits, 1947). Of course, the dangers of cross-cultural comparisons on a variable such as social responsibility, using a measure developed and standardized within one culture only, are obvious and great. Nevertheless, the sharpness of the differences obtained supports the possible fruitfulness of using relatively simple quantitative measures when gross differences between cultural groups are sought to test anthropological hypotheses.

The overall findings lend further support to the fruitfulness of using relatively simple preference measures for eliciting preferences for immediate, smaller as opposed to delayed, larger reinforcement in order to explore the dimension of delay of reinforcement and its antecedents and correlates. Before the relationships obtained can be generalized it will be necessary to test other cultural groups in similar experimental situations. The major significance of the present results is that they constitute an early step in testing empirically the popular theoretical conception, which has been relatively neglected experimentally, that delay of reinforcement is related positively to such concepts as "adjustment" and "maturity," and negatively to criminal, psychopathic, or socially irresponsible behavior. It remains for future research to isolate further the major component variable related to particular patterns of preference for delayed or immediate reinforcement.

SUMMARY

This experiment, using Trinidadian subjects aged 12 to 14, explored the relationship between preference for immediate, smaller reinforcement as opposed to delayed, larger reinforcement and (*a*) social responsibility as measured by an independently standardized scale, (*b*) delinquent behavior, and (*c*) accuracy in time statements. The primary results were (*a*) differences, as predicted, between delinquent and nondelinquent subjects, the former showing greater preference for immediate, smaller reinforcement, and (*b*) a tendency for subjects who preferred immediate, smaller reinforcement to show lower social responsibility scores, although in one of the two groups used this difference was significant only when restricted to subjects who showed consistent reinforcement preference patterns on three measures. Data relating delayed reinforcement preference patterns to accuracy in time statements were also presented. The overall findings suggest that quite simple and direct measures of reinforcement preference, in which smaller but immediately available rewards are pitted against larger but delayed rewards, may be effectively used to test often-discussed but rarely tested relationships between "maturity," "adjustment," "psychopathy," etc. and particular patterns of reinforcement preference with respect to delay in choice situations.

REFERENCES

EPLEY, D., & RICKS, D. Studies on time span. Harvard Psychological Clinic, 1959. (Mimeo)

FREUD, S. *Beyond the pleasure principle.* New York: Boni & Livewright, 1922.

HARRIS, D. B. A scale for measuring attitudes of social responsibility in children. *J. abnorm. soc. Psychol.*, 1957, **55,** 322–326.

HARRIS, D. B., CLARK, K. E., ROSE, A. M., & VALASEK, FRANCES. The measurement of responsibility in children. *Child Develpm.*, 1954, **25,** 21–28.

HERSKOVITS, M. J., & HERSKOVITS, FRANCES S. *Trinidad village.* New York: Knopf, 1947.

LEVINE, M., GLASS, H., & MELTZOFF, J. The inhibition process, Rorschach human movement responses, and intelligence. *J. consult. Psychol.*, 1957, **21,** 41–45.

LEVINE, M., SPIVACK, G., & WIGHT, B. The inhibition process, Rorschach human movement responses, and intelligence: Some further data. *J. consult. Psychol.*, 1959, **23,** 306–312.

MAHRER, A. R. The role of expectancy in delayed reinforcement. *J. exp. Psychol.*, 1956, **52,** 101–105.

MELIKIAN, L. Preference for delayed reinforcement: An experimental study among Palestinian Arab refugee children. *J. soc. Psychol.*, 1959, **50,** 81–86.

MELTZOFF, J., & LEVINE, M. The relationship between motor and cognitive inhibition. *J. consult. Psychol.*, 1954, **18,** 355–358.

MISCHEL, W. Preference for delayed reinforcement: An experimental study of a cultural observation. *J. abnorm. soc. Psychol.*, 1958, **56,** 57–61.

MOWRER, O. H., & ULLMAN, A. D. Time as a determinant in integrative learning. *Psychol. Rev.*, 1945, **52,** 61–90.

ROTTER, J. B. *Social learning and clinical psychology.* New York: Prentice-Hall, 1954.

SINGER, J. L. Delayed gratification and ego-development: Implications for clinical and experimental research. *J. consult. Psychol.*, 1955, **19,** 259–266.

SINGER, J. L., WILENSKY, H., & MCCRAVEN, VIVIAN G. Delaying capacity, fantasy and planning ability: A factorial study of some basic ego functions. *J. consult. Psychol.*, 1956, **20,** 375–383.

SPIVACK, G., LEVINE, M., & SPRIGLE, H. Intelligence test performance and the delay function of the ego. *J. consult. Psychol.*, 1959, **23,** 428–431.

(Received December 14, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 8-15

STABILITY AND FLUCTUATION IN THE POWER RELATIONSHIPS OF YOUNG CHILDREN[1]

ELIZABETH GELLERT[2]

Albert Einstein College of Medicine, Yeshiva University

THIS study explores one aspect of individual constancy in social interaction. In it, stability has been investigated in terms of day-by-day fluctuations in the dominant, acquiescent, and resistant behavior of preschool age children. For each subject, power related acts were recorded during three successive play sessions. Social and physical environments were held maximally constant by providing the same setting and an identical playmate for each child in all three play periods.

The stability of individual interaction patterns may be explored (*a*) by comparing the social behavior of the same subject (or subjects) in a variety of physiologic states, (*b*) by comparing the social behavior of the same subject in a variety of social situations, (*c*) by comparing the social behavior of the same subject in a variety of physical settings, and (*d*) by comparing the behavior of the same subject over time—that is, in terms of sequences or successive behavior samples collected for a more or less extended period. The present study is concerned with the last mentioned aspect of consistency.

A number of investigators have reported that persons tend to maintain consistent patterns of interaction in social situations (Bell & French, 1950; Hanfmann, 1935; Matarazzo, Saslow, & Guze, 1956; Matarazzo, Saslow, & Matarazzo, 1956; Merrill, 1954). However, as Chein et al. (1945) have pointed out, "... allowing the assumption of an overall stable pattern (of social behavior), the question of the average extent of day to day variability within this pattern of response remains relatively unexplored." If observations of social behavior are to be used in assessing personality, the degree of stability to be expected with regard to the behavior dimensions sampled becomes important. Particularly in the assessment of individual and group differences among the very young, measures are frequently based upon direct observation (Gellert, 1955). However, as Arrington (1943) has indicated, "the evidence contributed by time sampling and related studies (of children) concerning the stability of patterns of social . . . behavior observed to be characteristic of given individuals in particular situations is extremely meager." The present study represents an attempt to supply such evidence with respect to the power related behavior of preschool children.

METHOD

Subjects

The subjects were 16 boys and 16 girls enrolled at Harvard Preschool, ranging in age from 46 months to 62 months. All were Caucasian. Harvard Preschool is a nursery school attended, for the most part, by the children of college graduates. Most of the subjects had been acquainted with one another previous to taking part in the study.

Procedure

Stability in social behavior should be maximally apparent in a constant setting—that is, under conditions where there is little change from one time to another with respect to the physical and social environment wherein a person is observed. Such conditions were approximated in the experiment to be described. All experimental sessions took place in the same playroom. An identical array and arrangement of play materials was maintained. The toys provided were selected according to the following criteria: they should facilitate social interaction, they should lend themselves to a variety of uses so that the same subjects may play with them for repeated sessions without getting bored, they should appeal to preschool age boys and girls, they should not require adult direction for their use.

A trough filled with water and materials to be used for water play were provided; in another part of the

[1] The data cited here are part of a larger study entitled, "Patterns of Dominance, Submission, and Resistance in the Interaction of Young Children" (Gellert, 1956). This investigation was submitted in partial fulfillment of the EdD requirements at Harvard Graduate School of Education.

[2] The author is indebted to J. W. M. Whiting, John Carroll, and George Mandler for their guidance. Grateful acknowledgement is made, also, to the staff and children of the Harvard Preschool for their cooperation. Lelia Washburn assisted in the experiment. Support was also provided by NIMH Grant 2M 6418(C4) College 65, Interdisciplinary Program.

room, child-size furniture and housekeeping equipment were set out. An assortment of "dress up" clothes (donated adult wearing apparel) was included. In a third area, blocks, toy vehicles, and a train track set were placed. The playroom was equipped with one-way vision mirrors. Sound was transmitted to the adjoining observation booths.

The children were observed in pairs.[3] Each subject interacted with the *same* partner for three, 20-minute play sessions. Intervals between sessions, per pair, ranged from one day to a week. All pairs were homogeneous with respect to sex. Upon being taken to the playroom, the subjects were told that they could play with any of the things in the room. After demonstrating how she could be reached, if needed, the experimenter left the room, promising to return when playtime was over. For the most part, ongoing interaction was not interrupted by the experimenters. The subjects were not told that they were being observed. Individual behavior records of both children per dyad were charted concurrently. The experimenter and her assistant spot-checked their degree of agreement in recording their observations. Their average interobserver reliability was .81.[4]

Behavior was recorded in terms of *unit acts* defined as follows: Any scorable action or verbalization that is terminated by interruption as a result of other person's acts, by change of intent, change of direction, or by addition of new ideas on the part of the subject (that is, "Let's put on the shoes and then let's play house" is scored as two units because two ideas are expressed). Whenever a series of interacts turns into a ritualistic, repetitive game, only two units per subject are scored in a given category.

Selection and Development of Behavior Categories

Descriptions of power related behavior given by Murray (1938), Allport (1937), Cattell (1954), Chapple (Matarazzo, Saslow, & Matarazzo, 1956), Jack (1934), Anderson (1937, 1939), and Child (1954) were culled.[5] Actions or dispositions they considered to typify dominance or submissiveness were phrased in terms of apparently corresponding behavior observed in young children, and grouped into a set of discrete categories. During trial observations, these categories were modified and refined; several were deleted because corresponding acts occurred very rarely. (A complete list of the initial set of behavior categories appears elsewhere—Lambert, 1960.)

Definition of Dominance

In subsuming several behavior categories under one heading and by assigning the same value to acts in all categories, the categories are treated as though they were equivalent with respect to dominance. The behaviors represented by each are presumed to stem from a common inclination: the urge to dominate. While a significant correlation among dimensions of behavior assures neither equivalence nor common motivation, it does suggest that their association probably is not due to chance. With this consideration in mind, only those dimensions were retained in the final set of dominance categories that correlated significantly ($p \leq .05$) with composite dominance scores for each sex taken separately.

Composite scores were obtained by combining ranked performance in all 10 categories which initially were thought to reflect the intention to dominate. For each of these categories, the subjects were ranked in terms of frequency of acts. The sum of each child's rank positions on all 10 categories was taken as his *composite score*. Although positive Spearman rank correlations with the composite score were obtained for all the original categories, only 6 of them met the criterion for being retained. The final group of categories used to define dominance is described below.[6]

1. *Dominate play; instruct.*[7] P attempts to gain precedence in play; P pushes O out of the way; P changes what O has just built; P grabs toys O is using *in order to use them himself*; P initiates rules or refers to real or contrived "adult established rules" in order to direct play. (Only behavior occurring in the course of active, joint play was included in this category.)

   Examples:

   "If you want to play with me you'll have

[3] It was found that attending to more than two children at a time greatly reduced the reliability of observations.

[4] In addition to preliminary observations of individual children during play in their nursery school group, 12 pre-experimental paired play sessions were required to attain interobserver reliabilities between .75 and .85. The formula used to compute reliability coefficients was derived from Jersild and Markey (1935). It was:

$$\frac{2\,(\text{number of agreements})}{2\,(\text{number agreements}) + \text{A's omissions} + \text{B's omissions}}$$

(*Omissions* were units recorded by one observer, but not the other; *agreements* were the number of units, per category, that were recorded by both observers.)

[5] There is no assurance that the aggregate of behaviors and inclinations referred to in the literature under the denominations of dominance, ascendancy, leadership, and initiative spring from a common motivation. In fact, Clark and Birch (1945) have cited evidence that in the case of animals "priority of response in one type of competitive situation does not necessarily correlate highly with dominance in other competitive activities . . . " and that relative position in the pecking order may bear "little or no predictive relation to the initiation of new activities." Clearly, there is a continuing need to identify and circumscribe qualitative entities associated with power in social intercourse, and to study the interrelationships among such entities.

[6] The results of a more fundamental test of homogeneity, with respect to the categories used, are cited elsewhere (Lambert, 1960). This test was based upon data collected in a subsequent experiment.

[7] "P" refers to the subject whose behavior is being recorded. "O" refers to the other child comprising the dyad.

to . . ."; verbal threats to gain control, such as, "the teacher will be mad if you . . ."; also instructions such as, "I'll show you . . .," "you're supposed to . . .," "it goes this way," "we're going to . . .," "you better . . .," giving permission and assigning roles in play, such as, "you can be . . .," "it's my turn and then it's your turn . . .," "this can be yours," "you can be mother . . .," "nobody can come here but us"; refusing permission when asked, such as, "no, you can't have it."

2. *Positive mandate.* Spontaneous order or suggestion that does not emerge in the course of play. (Exception: orders that seem intended for attention seeking, such as, "look at me" or "look what I made.")

   Examples:

   "Come here!" "Let me take it!" "Put it down!"—when said aside from joint play.

3. *Negative mandate.* "Don't" "Stop!" etc. when used as an order rather than in the context of an altercation or as a defense against manifest domination on O's part. *Negation* has to be verbalized.

   Examples:

   Children are in different part of the play area and P attempts to control O's play by saying "Don't do . . ." "No, that's mine," etc.

4. *Suggest orientation.* This is a more gentle, or genteel form of domination than the above. Included are suggestions qualified by terms, such as, "please" "O.K.?" "would you like to . . ." or "let's . . . ." Conditional suggestions, such as, "would you do it . . ." "could you . . . ." (Exception: "Let me . . ." is scored as Positive mandate rather than in this category.)

5. *Boast.* Favorable comparison of self with other child, with regard to prowess, property, creation, appearance, etc.; boasting.

   Examples:

   "Mine is bigger" "I can jump way to the sky!" "My father can do it better than yours!"

6. *Tease.* P initiates a threatening gesture made "in fun," that is, without carrying out the implied threat; name calling; P derogates O, O's property, or O's efforts.

   Examples:

   "You didn't make it good." "Ha, ha, you can't catch me!" "Stinker!" "I don't like your . . . ."

Only behavior *initiated* by an individual, rather than occurring in response to domination on the part of a playmate, was defined as *dominance*. Behavior was scored in terms of apparent *intention* irrespective of its *outcome*. For example, *trying to appropriate another child's toy during joint play*, was scored as "Dominates play" (Category 1), whether the attempt was successful or not.

Definition of Submission

The subcategories comprising *submission* were selected by applying the same methods which have been described in relation to dominance. However, *responses* as well as *initiated* behavior reflecting deference were included in the definition of submission. The following categories were retained in the final set.

7. *Comply.* P complies in response to O's directions, domination, or bid for attention; P obeys or accepts O's impositions passively, desists, gives in; P concedes actively in response to O's domination; P does not defend self, property, or rights. (Act may be accompanied by verbal agreement, which is then *not* scored separately.)

8. *Agree verbally.* Verbal concession not accompanied by active compliance of any sort, in response to O's domination. (Exceptions: "yes" in response to O's asking permission; "yes" in response to O's asking a question.)

   Examples:

   "Yes" "O.K." "I will" "All right."

9. *Ask for permission, directions, orientation.* Almost any question is included here. (Exceptions: "What?" in response to not having understood something O said; "You know what I did? . . ." —which is really a comment designed to give information rather than obtain it.)

   Examples:

   "Can I put mine here?" "Where does this go?" "I don't know how to . . . ."

Definition of Resistance

With respect to the present subjects, it was found that opposition to impositions was not correlated significantly with dominance. Hence, such behavior was treated as a separate dimension, called *resistance*. The following categories constituted the subclassifications of resistance:

10. *Countermandate.* Domination *in response* to domination, noncompliance, or frustration occasioned by O. (Any scorable form of domination or noncompliance on O's part may provide a stimulus to a countermandate.) Conditional surrender, such as, "I will, *if* you . . ."; "I'll give it to you later"; counteraggression, counterteasing, counterboasting; sentences starting with "yes, but . . .," "no, but . . .," "but . . ."; counterarguments and disagreements in response to O's domination; a contrary order or a revocation of O's command. Included here is the mechanism of distracting O's attention as a defense—that is, O splashes water on P; P distracts him by saying, "Look what's out there!" (Exception: a simple, factual correction, when not said in the context of heated argument is scored in Category 1 rather than here.)

    Example:

    O: "Let's play with boats!" P: "No, let's build a castle."

11. *Noncompliance; self-defense.* Resistance to O's domination, per se—that is, not in conjunction with attempts to counterdominate. Persistence in the direction of own activity in the face of O's attempt to change it; defense of self, property, or rights. (Exception: when a refusal is followed by a counterattack or countersugges-

tion the whole episode is scored under Category 10.)

Examples:

"No!" "Don't!" "I don't want to . . ." "I won't"—in response to O's domination or in self-defense.

RESULTS AND DISCUSSION

Table 1 presents a summary of the subjects' scores in the experimental sessions. On the average, girls tallied more unit acts than did boys. Since this difference did not approach statistical significance (Median test: $p > .2$ —Edwards, 1954), the data for both sexes were combined in some procedures used for their analysis.

When performance in the three sessions was examined per individual, it was found that mean intersession differences in dominance ranged from 0 to 18 unit acts. While most children varied by less than seven dominance units when their records in three play periods were compared, nine subjects averaged session-to-session fluctuations of 12 to 18 unit acts. (Mean dominance per session, for boys and girls combined, was 19.7 units.) In terms of absolute frequency, then, some individuals varied considerably in dominance from period to period, whereas others demonstrated remarkable intersession stability. A comparable range of constancy was found with respect to submission and resistance.

In addition to the inspection of raw scores, the two following methods were used to determine the degree and manner of the children's consistency in power related behavior. First, the 30 subjects were compared with one another by ranking them according to the number of dominant acts recorded for each, in Sessions I, II, and III, respectively. In this way, three sets of ranks, each with an N of 30, were obtained. By using Kendall's coefficient of concordance (Siegel, 1956), the degree of association among these three ranks was assessed. The same procedure was applied to the records of submission and resistance. The results are presented in Table 2. They show that for the sample taken as a whole, significant coefficients of concordance were found with regard to dominance in all three sessions as well as in submission for all three sessions. However, resistance did not prove to be as markedly constant when ranks for the three sessions were compared.

A further measure of consistency is the extent to which an individual maintains the position of greater (or lesser) dominance, submission, or resistance relative to his partner. In

TABLE 1

AVERAGES AND VARIABILITIES OF SCORES

(N = 14 Boys + 16 Girls)

	Session I			Session II			Session III			All Sessions		
	Mdn	*M*	*SD*	Mdn	*M*	*SD*	Mdn	*M*	*SD*	Mean per Session	Range	Mean Difference[a]
Dominance												
Boys	11.5	15.4	10.3	13.5	19.0	13.2	11.3	17.2	14.4	17.2	0–23	6.7
Girls	17.5	20.0	13.8	20.0	24.2	13.2	19.5	22.2	12.4	22.1	5–29	7.8
Submission												
Boys	11.5	11.9	6.6	12.5	14.2	9.4	13.0	14.4	7.7	13.1	1–43	6.1
Girls	15.0	17.8	13.1	18.0	20.3	10.7	19.5	20.8	12.8	19.6	2–47	7.7
Resistance												
Boys	5.5	6.8	4.3	8.5	8.9	5.9	8.0	8.9	5.5	8.7	1–24	5.8
Girls	8.0	8.3	5.1	15.0	14.6	6.2	16.5	15.4	9.4	12.8	1–33	7.8

[a] For each subject, the absolute differences in scores between Sessions I and II, Sessions II and III, and Sessions I and III were summed and divided by 3 in order to obtain his Mean Difference score. The numbers in this column represent means of these scores.

TABLE 2
COEFFICIENTS OF CONCORDANCE FOR THE THREE PLAY SESSIONS
($N = 30$)

	W	p^a	r^b
Dominance			
Category 1 Dominate Play	.65	$< .01$	.48
Categories 2 and 3 Mandate (positive and negative)	.75	$< .001$	.63
Category 4 Suggest Orientation	.59	$< .05$	.38
Combined Score	.78	$< .001$	.67
Submission			
Categories 7 and 8 Comply and Agree verbally	.67	$< .001$	.51
Category 9 Ask for permission, directions, orientation	.73	$< .001$	.56
Combined Score	.73	$< .001$	.56
Resistance			
Category 10 Countermandate	.55	$< .02$	.33
Category 11 Noncompliance, self-defense	.43	$> .05$ *ns*	.14
Combined Score	.49	$< .10$ *ns*	.23

Note.—Ws for Categories 5 and 6 were not computed separately because of the generally low frequency, per subject, of behavior assigned to these classifications. The Combined Score is the sum of entries in all categories per behavior system.

[a] One-tailed test.

[b] Average value of the rank correlation coefficients (Edwards, 1954).

order to determine this, the records of the subjects comprising each dyad were compared with one another. Figure 1 indicates which subject had the higher dominance, submission, and resistance score, respectively, in each of the sessions.

It is apparent that there was a marked tendency for the same individual to tally the higher dominance score during all three sessions. This also held true with respect to submissive acts. For two-thirds of the pairs, there was no shift in dominance balance when performance in three sessions was compared. Thus, the same individual maintained the higher dominance score throughout in a significant proportion of cases ($p < .001$). In 8 out of 14 pairs, the submission balance remained the same for all three play periods ($p < .01$).

While most pairs maintained consistent relationships with regard to dominance and submission, the tendency to resist imposition proved to have no greater than chance stability when it was viewed in terms of preponderance of tallies within each dyad, during the three play sessions. The apparent difference in stability between resistance on the one hand, and dominance and submission on the other, may have resulted from any of the following factors.

Resistant behavior may have been recorded with less accuracy than dominance or submission. It will be remembered that this dimension initially was not viewed or treated as a separate behavior system. Therefore, the observation categories comprising the system, *resistance*, may have been comparatively less exhaustive than those classified under dominance and submission. Thus, instances of expressing resistance in ways not included in the present category system may have remained unrecorded. Such omissions may account for the apparent inconsistency of resistant behavior.

Perhaps more than three 20-minute play sessions are required for the establishment of a stable resistance level per individual. Thus far, it has been assumed that the experience of interacting with the same partner in the same room for a series of play periods represented a repetition of rather equivalent situations for most subjects. This is not necessarily the case. For example, individual variability in resistance, when performance in three sessions was compared, might have been a function of systematic differences in conditions—perhaps due to phase effects—among the three periods. In order to examine the probability that the behavior samples comprising the three sessions were drawn from an identical population of sessions, the Friedman two-way analysis of variance by ranks (Siegel, 1956) was computed for each of the power dimensions studied. No significant intersession differences were found for dominance or submission. However, it was dicovered that the series of play sessions probably did not represent equivalent situations with respect to the

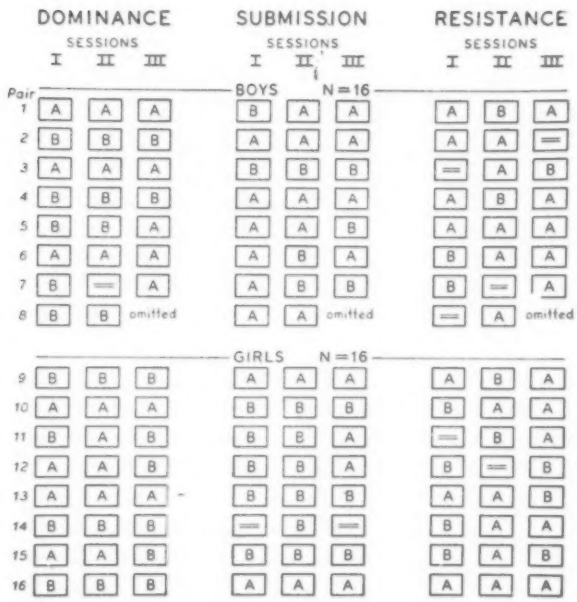
DOMINANCE
SUBMISSION
RESISTANCE
SESSIONS
I II III
Pair
BOYS N=16
GIRLS N=16
omitted
Explanation: The letters, A and B refer to the two children in any dyad. The letter in each box indicates which subject tallied the higher score per session, for each dimension. "=" refers to tied scores.

FIG. 1. Maintenance of higher dominance, submission, and resistance score for members of each dyad.

expression of resistant behavior ($p < .02$). For most subjects, resistance was highest in the second play period, perhaps reflecting an ongoing power struggle. Such an interpretation is suggested by the findings of Heinecke and Bales (1953) who reported that, among groups of adults, a status struggle occurred in the second session of any series of group meetings. In exploring the vicissitudes of power in small groups, it is desirable to use subjects who were unacquainted with one another previous to being observed in interaction—as was done by Bell and French (1950). Since the present subjects attended the same nursery school, they had opportunities to establish power patterns with respect to one another prior to the experiment. This factor may have interfered with the development of more pronounced indications of a status struggle during Session II. A more extended series of sessions would be required to explore the possibility that periodic power struggles may take place in the course of repeated association in any stable, small collection of individuals. Such an investigation would have obvious relevance to the study of interaction patterns in the nuclear family and in other constant primary groups.

Some dimensions of social interaction probably are highly innerdetermined, whereas others are primarily situation-determined. If social behavior were arranged along a continuum in such terms, dominance and submission might well fall closer to the "innerdetermined" end than would resistance. Such a view of resistant behavior is consistent with the findings reported in Caille's (1933) study of resistant behavior in preschool children.

It is tempting to speculate about the possible effect of developmental level upon the stability of interaction patterns. The assumption is often made that "the personality achieves a type of inner organization which becomes more and more powerful with the passage of time, and which frees the personality to a greater degree from the dominance of the situation" (Anderson, 1948). In other words, as he grows older, the individual is thought to become increasingly consolidated, or fixed, with respect

to his characteristic patterns of behavior. Even if this were so, all aspects of social behavior probably would not mature at the same rate. Hence, it may be that for most individuals, characteristic dispositions with respect to dominance and submission are well established by age 5, while their inclination to exhibit resistance has not yet achieved a commensurate degree of consolidation.

Arrington (1943) has suggested that consistency may be characteristic of some individuals, and inconsistency of others. The data were examined in search of such pervasive differences in intrapersonal stability. This was done by comparing each subject's average difference scores (see Table 1 for derivation of these scores) for dominance, submission, and resistance. No coincidence was found in degree of individual variability with respect to these three dimensions. Hence, the data do not support Arrington's proposition. However, the evidence with regard to this point must be viewed with considerable reservation, since the present study provided a very limited measure of individual patterns of consistency. A better picture could have been obtained if the subjects had been observed for a longer series of sessions.

The method of recording behavior used in this investigation was confined to measuring the frequency of power related acts. Intensity, duration, and latency of relevant behaviors were not assessed. Neither were group differences in interaction rate taken into account. The difficulties of devising and putting into operation a procedure that includes these aspects of social behavior led to the adoption of the present scheme of assessment. Despite the obvious limitations of frequency counts in comparing samples of behavior, the data suggest that such counts may have some merit as an index of individual consistency with respect to some of the dimensions involved in the present study.

SUMMARY

This study was designed to explore day-by-day fluctuations in the power related behavior of young children. Sixteen pairs of children were observed in a constant setting during three, 20-minute play periods. Frequency of dominant, submissive, and resistant acts was recorded for each individual.

For many subjects, a fairly high degree of intersession variability was found when performance during the three play periods was compared. However, for the group as a whole, interindividual differences exceeded individual inconsistency with respect to dominance and submission. The majority maintained fairly stable rank positions within the group in terms of these dimensions during the three sessions. In a significant proportion of dyads, the same child had the higher dominance and submission scores, respectively, during all three play periods, when the two playmates' records were compared. Such stable individual behavior patterns were not found with respect to the expression of resistance.

REFERENCES

ALLPORT, G. *Personality: A psychological interpretation.* New York: Holt, 1937.

ANDERSON, H. H. Domination and integration in the social behavior of young children in an experimental play situation. *Genet. psychol. Monogr.*, 1937, **19**, 341–408.

ANDERSON, H. H. Domination and integration in the social behavior of kindergarten children. *J. exp. Educ.*, 1939, **8**, 123–131.

ANDERSON, J. E. Personality organization in children. *Amer. Psychologist*, 1948, **3**, 409–416.

ARRINGTON, RUTH E. Time sampling studies of social behavior. *Psychol. Bull.*, 1943, **40**, 81–124.

BELL, G. B., & FRENCH, R. L. Consistency of individual leadership position in small groups of varying membership. *J. abnorm. soc. Psychol.*, 1950, **45**, 764–767.

CAILLE, R. K. Resistant behavior in preschool children. *Child develpm. Monogr.*, 1933, No. 11.

CATTELL, R. B. *Description and measurement of personality.* New York: World Book, 1954.

CHEIN, I., et al. Consistency and inconsistency in intergroup relations. *J. soc. Issues*, 1945, **5**(3).

CHILD, I. L. Socialization. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge, Mass: Addison-Wesley, 1954. Pp. 655–692.

CLARK, G., & BIRCH, H. G. Hormonal modifications of social behavior: I. The effect of sex-hormone administration on the social status of a male castrate chimpanzee. *Psychosom. Med.*, 1945, **7**, 321–329.

EDWARDS, A. L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954.

GELLERT, ELIZABETH. Systematic observation: A method of child study. *Harv. educ. Rev.*, 1955, **15**, 179–195.

GELLERT, ELIZABETH. Patterns of dominance, submission, and resistance in the interaction of young children. Unpublished doctoral dissertation, Graduate School of Education, Harvard University, 1956.

HANFMANN, EUGENIA. Social structure of a group of kindergarten children. *Amer. J. Orthopsychiat.*, 1935, **5**, 407–412.

HEINICKE, C., & BALES, R. F. Developmental trends in the structure of small groups. *Sociometry*, 1953, **16**, 17–39.

JACK, LOIS M. An experimental study of ascendant behavior in preschool children. *U. Iowa Stud., Stud. Child Welfare*, 1934, **9**, 7–65.

JERSILD, A. T., & MARKEY, F. V. Conflicts between preschool children. *Child develpm. Monogr.*, 1935, No. 21.

LAMBERT, W. W. Interpersonal behavior. In P. H. Mussen (Ed.), *Handbook of research methods in child development*. New York: Wiley, 1960. Pp. 854–917.

MATARAZZO, J. D., SASLOW, G., & GUZE, S. B. Stability of interaction patterns during interviews: A replication. *J. consult. Psychol.*, 1956, **20**, 267–274.

MATARAZZO, J. D., SASLOW, G., & MATARAZZO, R. The Interaction Chronograph as an instrument for objective measurement of interaction patterns during interviews. *J. Psychol.*, 1956, **41**, 347–367.

MERRILL, BARBARA. A measurement of mother-child interaction. In E. W. Martin & Celia B. Stendler (Eds.), *Readings in child development*. New York: Harcourt, Brace, 1954. Pp. 346–357.

MURRAY, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

(Early publication received September 6, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 16-23

THE PREDILECTION FOR SINGLE ORDERINGS[1]

CLINTON B. DE SOTO[2]
Johns Hopkins University

PEOPLE are always ready to rate one another. But their ratings have a certain defect. They tend to correlate unduly. Rated voice quality is likely to correlate very highly with rated intelligence, even if they are in fact quite independent traits. Thorndike (1920) christened this tendency "halo." Since then it has been one of psychology's bread-and-butter phenomena. It is a strong, impressive phenomenon, little known to the public, yet easy to describe, and controllable in some degree with devices like forced-choice ratings. But Thorndike's act seems to have taken the surprise out of it, so that it has remained for 40 years mainly a bother for the practical psychologist, rather than a poser for scientific psychology. So little regard has been given to explaining it, in fact, that the best explanatory discussion Guilford (1954) finds to cite is that published by Symonds (1925). And this is not because Symonds' reasons for halo are definitive—indeed, they seem to suffer mutual inconsistency.

The classic view of halo is that it shows a reliance on general impressions arising from stupidity or ignorance or carelessness. In support of this view is Symonds' assertion that halo is especially strong for ratings on traits that are not commonly observed or thought about. People are especially stupid and careless in handling these unfamiliar, unimportant traits.

But the argument that halo is merely a puerile failing runs into complications with other assertions by Symonds. He says that halo is also especially strong for traits of high moral importance and for traits involving reactions with other people. Why should people be especially stupid and careless in dealing with such traits? Still more difficulty for the argument is provided by Sherif, White, and Harvey (1955), who found that even on a task of supreme importance to them, handball throwing, boys based their predictions and judgments of another's performance mainly on his status in the group. It has in fact been suggested by some that halo is stronger for traits that are important to people than for those that are unimportant (Asch, 1952, p. 230; Riecken & Homans, 1954).

At first glance, it might seem that these complications make halo a more subtle and important phenomenon than the general-impression hypothesis would have it, but an augmented general-impression hypothesis can triumph over them. It is argued that in these cases the affective reaction toward a person governs judgments of him, preventing differentiated judgments. It might seem surprising that emotion should consistently override thought. It might even be supposed that the affective reaction itself could be differentiated, permitting differentiated judgments. But no serious doubts have been expressed about this view of halo, and it remains the commonly accepted one, with little elaboration and less competition.

There is, however, another kind of difficulty which is harder to incorporate under the general-impression hypothesis: people have a positive predilection for the real-world counterpart of halo, for a high correlation between the variables on which people are ranked. A number of studies, mostly by sociologists, demonstrate this fact. Lenski (1954) has shown that if a person has discrepant scores on the status variables of society at large (income, occupation, education, ethnic background), he tends to be dissatisfied. Adams (1953) found that the morale of bomber crews was positively correlated with the average intercorrelation of their status variables, rank, flying time, education, reputed ability, popularity, length of service, combat time, and position importance. Homans (1953) showed that discrepancies between pay and prestige of office jobs produced dissatisfaction. Exline and Ziller (1959) found that congeniality in small groups depended on the correlation of the members' ability and voting power, which were controlled experimentally. And Whyte (1943), in his

[1] This work was supported in part by Grant NSF-G10884 from the National Science Foundation.

[2] The writer is grateful to Sergei V. Timkovsky for aid in the collection of data.

famous study of a street-corner gang, the Nortons, describes how the members of this gang actively made bowling scores correlate with status in the group, by appropriate encouragement and heckling. Indeed, Benoit-Smullyan (1944) argues that what produces discontent and rebelliousness in people, whether in a small group or great society, is not so much oppression or low status itself, as it is the possession of disparate statuses. The rebel isn't the man who is low on all ladders; he is the man who is low on some and high on others.

It is hard for the general-impression hypothesis to cope with these cousins of the halo effect. The general-impression hypothesis can directly explain only why people misperceive, not why they might be dissatisfied with what they perceive. It is possible that in some cases misperception plays a mediating role, leading to inappropriate behavior which results in punishment and dissatisfaction with the social structure. But in many cases the discrepant orderings are clear enough and misperception unlikely.

It seems indeed that people have a predilection for single orderings, an antipathy toward multiple orderings. This is not to say that the predilection itself is not rooted in the relative ease with which people can handle single orderings as compared with multiple ones. In recent years psychologists have realized that people find many ways and go to surprising lengths to avoid cognitive strain (Festinger, 1957; Heider, 1958). But there is a big difference between the trivial statement that multiple orderings are naturally harder than single orderings and the statement that they are so much harder that people are discontent with them and actively seek to transform them to single orderings.

Since social scientists are people first of all, it might be predicted that they bear the same predilection. Do they? It appears that they do and that the predilection is a prime source of difficulty for social theory. Benoit-Smullyan (1944), Lenski (1954), and Rose (1958) all have chided their fellow social scientists for their tendency to treat society as having a single hierarchy or ordering rather than distinguishing the ordering variables of wealth, power, education, prestige, and so on. Lenski says that from Aristotle on, most social philosophers and social scientists have made this particular oversimplification. Despite recognition early in this century of the desirability of considering these ordering variables separately, there is a stubborn, recurring urge to reduce them to a single ordering. The modern, sophisticated way to do it is to average each subject's ranks on the several orderings, an involved procedure which loses much and gains nothing, according to Rose and Lenski. Another way, one discussed by Benoit-Smullyan, is to insist that only one of the ordering variables is important, the others being only direct or indirect expressions or consequences of it. Thus a Marxist might say that what counts in a Western society is wealth, and power and prestige are merely derivatives of wealth. Pareto might say instead that wealth is only the material expression of social power.

In treating societies and cultures and social systems themselves, social scientists have also suffered from the predilection for single orderings. The early sociologists and cultural anthropologists taught, for example, that there is a single ordering of cultures from primitive to advanced—the so-called doctrine of unilinear evolution (Steward, 1953). This teaching is rejected by most anthropologists today. What is wrong with it? Not that there are no ways in which cultures are ordered from primitive to modern, but that there are many. A given culture might occupy a high rank in the ordering by metallurgy but a low rank in the ordering by mathematics. The early anthropologists blinded themselves to this multiplicity of discrepant orderings in their predilection for a single ordering. Even today some anthropologists cling to the conception of cultures as having a single ordering by reducing the number of steps in the ordering to a paltry three or four—"savagery," "barbarianism," "civilization." This degradation of the doctrine keeps it alive for the sake of the predilection, but in a form that is as useless as it is unassailable (Steward, 1953).

Still another topic on which thinking may be blurred by the predilection for single orderings is leadership. Scientists and laymen both call a number of different attributes and functions leadership—popularity, power, skill, prominence, for example. All of these are variables which to some extent order the members of a group, and in principle it is quite possible for

these orderings to be discrepant, with different members at the top. But of course people will expect all the orderings to be the same, will expect the most popular member to be the most powerful and most skillful and most prominent, to be, in a word, the leader. It seems so fitting to say, "First in war, first in peace, and first in the hearts of his countrymen." This expectation lands the social scientist in a curious impasse. To begin with, he has the predilection, like anyone else, and his discussion of leadership is likely to show it. But at the same time as a wiser colleague criticizes him for confusing popularity with eminence and skill with influence, there may be justification for him in the fact that his subjects, having the predilection too, tend actually to bring about the single ordering, investing a single leader with all these things. In fact, if the leader renounces some of his leadership functions and tries to distribute them among his followers, as many enlightened social scientists would have him do, what is his reward? He has violated his followers' expectation of a single ordering and they are likely to be displeased rather than grateful (Berkowitz, 1953).

Even in colloquial language one can detect people's tendency to act as though there is only one ordering of a set of people. They seem to delight in gross expressions like "Joe's tops," or "Joe's the greatest," avoiding any hint that there may be orderings on which Joe does not stand uppermost.

Probably in any case where people are made to think of an ordering, even an ordering of elements other than people, they will resist thinking of any disparate ordering. If so, a kind of halo effect may be a more pervasive problem for judgments and ratings than is ordinarily realized. The study of social desirability of traits and behaviors seems to have run afoul of this problem. At first, it would seem a straightforward matter to have the subjects judge items drawn from personality inventories as to their social desirability (Edwards, 1957). But it does not turn out to be straightforward (De Soto, Kuethe, & Bosley, 1959; Johnson, 1955). These items were originally selected so as to vary widely on desirability for personal well-being or health, and this variation seems to provide a salient ordering of the items for the subjects. As a consequence, they have a troublesome tendency to judge their desirability for personal well-being when they are supposed to judge their social desirability.

Even in defining the key mathematical property of an ordering relation called transitivity, one can encounter people's infelicity with disparate orderings. The requirement for transitivity of a relation is usually stated something like this: If A bears the relation to B, and B bears the relation to C, then A must bear the relation to C. It is sometimes convenient, however, to state it thus (De Soto & Kuethe, 1959): If A bears the relation to C, and C bears the relation to B, then A must bear the relation to B. These two formulations are equivalent, mathematically speaking. But people find the former much easier to apprehend than the latter, which seems to confuse if not mislead them. It asks them to think of the ordering $A \rightarrow C \rightarrow B$, a task that would be easy but for the obtrusion of the overlearned disparate ordering $A \rightarrow B \rightarrow C$. In the other formulation, of course, $A \rightarrow B \rightarrow C$ is the only ordering that presents itself, and understanding is quick.

EXPLANATION OF THE PREDILECTION: AN EXPERIMENT

Evidently the predilection for single orderings materializes in many forms. Knowing this should eliminate the surprise and bewilderment of people like Rose (1958) who encounter it unawares. But what more can be said about it? Can one go beyond it to its source or explanation? Certainly the data cited so far help little in the quest; they simply show its generality. But how can one do more? Can one even pose some possible sources or explanations?

It has been suggested that multiple orderings are odious because they are difficult. Is there any way to measure this difficulty? Is there any way to demonstrate that multiple orderings are more or less difficult than something else?

The difficulty of single social structures, including single orderings, has been successfully measured by the time people take to learn the structure under controlled conditions (De Soto, 1960). It should be possible to apply this measure to multiple orderings. It should take a long time to learn two orderings of a set of people. But a long time in comparison

with what? Certainly it should take longer to learn two orderings of a set of people than one ordering of them, if only because of the difference in size of the tasks. It would be more enlightening to compare the learning of the two orderings with other tasks of comparable size. Comparison of this sort might decide whether or not there is something peculiarly difficult about multiple orderings, and if there is, what it is.

The following experiment provides such comparisons. In this experiment all subjects were given tasks of about the same magnitude: all were required to learn two social structures or organizations at once. In one condition subjects were required to learn two discrepant orderings of a set of names: an ordering by height and an ordering by age, the two orderings being uncorrelated. Subjects in a second condition had to learn two discrepant groupings of a set of names: a grouping by majors and a grouping by houses—a division of the names into groups according to majors and a disparate division of them into groups according to houses. Subjects in a third condition were required to learn one ordering and one (uncorrelated) grouping of a set of names, for example, an ordering by height and a grouping by majors. In a fourth condition subjects were required to learn an ordering by age of one set of names and an ordering by height of another set of names—two orderings, as in the first condition, but of distinct sets rather than the same set. In a fifth condition subjects learned a grouping by majors of one set of names and a grouping by houses of another set. In a final condition subjects learned an ordering of one set of names and a grouping of another set. In the conditions with one set of names, only men's names were used; in the conditions with two sets of names, one set was composed of men's names and one of women's names, to make them distinct. In all cases a grouping of a set was a division of them into two groups.

The experimental conditions, as well as one set of experimental results, are summarized in Table 1. The conditions were chosen with a view to testing several possible explanations of the predilection for single orderings.

1. If every condition in the left-hand column should turn out to be harder than its counterpart in the right-hand column, one would have evidence that it is peculiarly difficult to handle two organizations of the same set, and the difficulty of two orderings of the same set would emerge as a special case of this general statement.

2. If the upper conditions should turn out to be harder than the bottom condition of each column, indicating that people had difficulty learning two orderings at once or two groupings at once, whether they were dealing with one or two sets of names, one would be led to another general statement: it is difficult to handle two organizations of the same kind at the same time. Again, the difficulty of two orderings of the same set would emerge as a special case.

3. It would be quite conceivable, of course, for both of the preceding possibilities to be true, in which case the upper conditions of the left-hand column should be hardest of all and the lowest condition of the right-hand column easiest, with the other conditions intermediate in difficulty. In this case, the halo effect and its cousins would be the resultant of the two factors acting in concert. People have difficulty in handling two organizations of the same set of objects; people have difficulty in handling two organizations of the same kind; the difficulty of two orderings of the same set consists of these difficulties combined.

4. On the other hand, if the two upper conditions of the left-hand column should emerge as the hardest of all but the bottom condition of the right-hand column should not turn out to be the easiest of all, the preceding general statements would be ruled out. Organizations of one set would be harder to learn than organizations of two sets only if they were organizations of the same kind. Organizations of the same kind would be harder to learn than organizations of different kinds only if they were organizations of the same set. The presumptive general statement would be: it is hard to handle two organizations of the same set of the same kind.

5. If the top condition of the left-hand column alone turned out to be abnormally difficult, any general statement would be barred. The conclusion would be that there is something uniquely difficult about two orderings of the same set, a difficulty not paralleled elsewhere and therefore not derivable from a general principle as a special case.

6. Finally, if the conditions emerged as equally difficult, one would be led to conclude that there is after all nothing peculiarly difficult about discrepant orderings of a set, that other arrangements of equal proportions or complexity are equally difficult.

Method

To check the generality of whatever pattern of difficulties emerged, the experiment was performed with two different arrangements of the subject's task. In the first version of the experiment, results for which are reported in Table 1, the subjects were taught the social structures in an expedient way: the experimenter recited the structures to subject on each trial, then questioned him about the relation between pairs of people. A typical question would be, "Is Bob older than Dave?" After obtaining the subject's answer, the experimenter gave him the correct answer. In the second version (Table 2) the experimenter was niggardly with information about the structures: he questioned the subject about the relation between pairs of people and (after obtaining the subject's answer) gave the correct answer to each question, but gave no other information. This technique is the one used earlier for the study of learning a single social structure (De Soto, 1960).

In the first version of the experiment the social structures had six members each; in the second version, only three, because learning of larger structures would have been too slow.

In the first version, a trial consisted of the experimenter's recitation of the social structure followed by 6 questions, 3 about each structure. In the second version, a trial consisted of 12 questions, 6 about each structure. For both versions, the criterion of mastery was two successive trials on which the subject answered all questions correctly.

The subjects for the experiment were male undergraduates of Johns Hopkins University. Each subject was run individually and served in a single experimental condition. Five subjects were used in each condition for the first version; 16 in each condition for the second version, which produced greater individual differences.

TABLE 1
MEAN TRIALS TO CRITERION
(First Version of Experiment)

	One Set	Two Sets
Two Orderings	16.2	6.6
Two Groupings	10.0	6.6
Ordering and Grouping	7.2	7.0

TABLE 2
MEAN TRIALS TO CRITERION
(Second Version of Experiment)

	One Set	Two Sets
Two Orderings	15.7	11.5
Two Groupings	9.7	6.6
Ordering and Grouping	8.4	10.5

Results

Examination of Table 1 shows that the subjects chose Outcome 4. Table 2 is a little more complicated, but seems to show the same thing. The interaction pattern is the same, as can be seen more easily if the middle row of figures is increased by an additive or multiplicative constant. There may be justification for doing this in the possibility that a grouping is easier than an ordering for three-member sets although not for six-member sets. Certainly the results in Table 2 are in far better correspondence with Outcome 4 than with any of the other proposed outcomes. An analysis of variance performed on the Table 1 data yielded an interaction significant at the .05 level; another performed on the Table 2 data yielded an interaction significant at the .10 level; and one performed on the pooled data yielded an interaction significant at the .01 level.

The experiment then yields as a strong candidate for the explanation of halo the statement: it is hard to handle two organizations of the same set of the same kind. But, as always, it seems there must be an explanation lurking behind the explanation, an explanation of the explanation. Why is it hard to handle two organizations of the same set of the same kind?

OPERATION OF SCHEMATA

So far, it has been taken for granted that people are facile in treating a set of objects as having a single ordering or a single grouping. Not that there seems reason to doubt the facility—it is manifest in many ways. De Soto and Kuethe (1958), for example, found that children as young as 9 can usually answer questions about the mathematical properties of ordering and grouping relations. They know that if A is older than B, it follows that B is not older than A (asymmetry). They know that if A is older than B, and B is older than C, it follows that A is older than C (transitivity). They know that if A is in the same grade as B, it follows that B is in the same grade as A (symmetry). They know that if A is in the same grade as B, and B is in the same grade as C, it follows that A is in the same grade as C (transitivity). Another kind of evidence

for the facility is provided by Scott's (1956) study of the errors made by enlisted submarine personnel in naming people in their organizational chart. Scott found that omissions (failing to recall names) were the most common kind of errors. Echelon errors (placing people too high or too low in the chain of command) and unit errors (placing them in the wrong group) were relatively rare. Recalling names gave the subjects the most difficulty, not placing them in the ordering or grouping once recalled.

Clearly, there is no mistake in saying people are facile in handling single orderings or single groupings. But there may be a mistake in leaving it at that. For people are not so good at everything. Valuable perspective on the ease with which people answer the questions about mathematical properties of ordering and grouping relations can be gained by contrasting it with the difficulty they have with other formally similar questions. They have extraordinary difficulty with simple questions about logic, for example. If you tell someone that A implies B, and B is true, he is likely to infer mistakenly that A must be true. Even a slight variation of the relation "is older than," changing it from a simple ordering to a weak ordering, produces dramatic difficulties. If you tell people that A is at least as old as B, and ask if B is at least as old as A, many will say "yes," others will say "no," and only a minority will correctly say he may or may not be (De Soto & Kuethe, 1958).

Another contrast with people's skill with orderings is their ineptitude with dominance structures (De Soto, 1960). Dominance structures are poor relatives of orderings, having two of the three properties of an ordering—asymmetry and completeness, but not transitivity. They are called dominance structures because social dominance relations among animals (especially peck rights among birds) produce such structures. They are primitive and simple social structures, the inevitable outcome of the pair-wise encounters among animals by which they are formed. Yet most people, scientists included, consider them puzzling and complex as compared with orderings. Learning such a structure takes much longer than learning an ordering; people fail to profit from the properties it does have—in fact, they may even be hindered by them, especially if they expect an ordering.

In short, people are very good at handling orderings and groupings but very bad at some other similar tasks. It can be inferred that their facility with orderings and groupings does not reveal any general power of thought, but rather quite specific thinking mechanisms or processes which might aptly be called schemata. Thus one can say that people pretty universally have well-functioning ordering schemata and grouping schemata. This proposal may smack of the instinct-naming or faculty-naming approaches which are now anathemas in psychology, but it need not mark the end of investigation, as naming an instinct or faculty so often did.

For present purposes, it has value in suggesting a different tack in the explanation of the predilection for single orderings. Like most human traits and functions, a schema surely has adaptive or survival value. Skirting teleology, as biologists are prone to do, one can ask, what is a schema for? How does it serve the organism? What properties should a good schema have?

Although it has rarely been made explicit in discussions of schemata, it appears that buried in the traditional meaning of schema are two key requirements. A schema contributes to cognition, but it only contributes. An ordering schema, for example, should enable a person to treat objects as ordered, but should not thereby blind him to other urgent aspects of the objects. In everyone's life there must be countless situations where the ability to apply two schemata to a set of objects is well rewarded. This ability would develop through evolution or learning or both—and is evidenced in the lower left-hand mean of Tables 1 and 2. The ability to apply the same schema twice to a set of objects, on the other hand, might be a special case, less often needed, less often rewarded, hence less developed. The other key requirement of a schema is that it is multiply applicable. Treating one set of objects as ordered must not bind the ordering schema so as to prevent treating another set as ordered. Again, the development of this ability might not be accompanied by equal ability in the special case of multiple applications of the same schema to the same

set of objects. Thus the upper means of the left-hand columns are greater than the upper means of the right-hand columns.

This approach to the problem keeps in the forefront the fact that it is an achievement to treat objects as ordered at all, that the successes have to be accounted for as well as the failures. At the same time it exposes our crippling ignorance of the dynamics of schema arousal and application.

In the enormous literature on serial learning, for example, one finds almost no discussion of the instructions with which the subject is initially induced to treat the items as ordered. Why this nonconcern with the instructions, almost a denial that they have been used? The explanation probably lies in the faith that all that goes on in such learning is conditioning, a primitive process that shouldn't require instructions. But the fact is that the instructions are critical. If the subject is told to remember the items in order he does so much better than if he is simply told to remember the items (Horowitz, 1960; Kay & Poulton, 1951; Postman, 1956). Appropriate instructions evidently arouse a helpful ordering schema in the subject—but this arousal is accomplished so easily and quickly that it is easy to overlook that it has occurred at all.

Likewise in the literature on halo there are only suggestive glimpses of the ordering schema being engaged and disengaged. It has been noticed that the halo effect is less when judges rate one trait at a time than when they rate one person at a time (Guilford, 1954). Why might this be? A plausible explanation is that two applications of the ordering schema to a set of objects conflict only if they occur simultaneously, that is, if the subject is thinking of both orderings at the same time. When a person judges on one trait at a time, he has to think of only one ordering of the ratees at a time, whereas when he judges one person at a time, he has to think all at once of the many orderings on which the ratee occupies positions. This explanation would also serve for another phenomenon of ratings—the tendency for nearby traits on a rating form to intercorrelate more highly than remote ones. This tendency, only recently discovered, has been named the proximity error (Guilford, 1954), but it might be regarded as showing simply the gradual dying away or disengaging of earlier applications of the ordering schema. The proposition that the halo effect depends on the subject's thinking of the orderings involved simultaneously, disappearing to the extent he is able to think of them one at a time, may even help explain the rather puzzling observation noted earlier that halo may be stronger for orderings that are important to people than for the unimportant ones. The argument is this: the more important orderings are to a person, the more time he will spend thinking about them, and as a statistical consequence, the more likely he is to think about them simultaneously.

But it is easy for discussion of schemata and their operation to get far out of empirical bounds. There seems to be good evidence that people have a far-ranging predilection for single orderings. There seems to be evidence that the predilection and other phenomena can be elucidated by crediting people with an ordering schema that is subject to certain principles of operation. Rather than more speculation, what is needed now are experimental studies frankly concerned with these things.

SUMMARY

It was proposed, as an inductive generalization, that people have a predilection for a single ordering or ranking of a set of people or other objects, an aversion to discrepant orderings of the set. In ratings, this predilection shows up as the halo effect—the tendency to reduce discrepant orderings on different traits to a single ordering by merit. In social structures, including small groups and large societies, it shows up in general dissatisfaction with discrepant orderings of the members and in efforts to eliminate the discrepancies. In the theorizings of social scientists about society and culture it shows up as a stubborn urge somehow to reduce discrepant orderings of people, or classes, or cultures, to single orderings. And in other judgmental and intellectual tasks people show difficulty in handling discrepant orderings which contrasts strongly with their facility in handling single orderings.

Once the generalization had been established by appeal to such phenomena, an effort was made to give it context and explanation

through experimentation. The difficulty of learning two discrepant orderings of a set of people was compared with that for five other tasks in which subjects also learned two social structures at once. It was found that two orderings of a set of people were relatively difficult to learn, providing one more instance of the generalization. From the pattern of difficulties over the six conditions, however, it was inferred that the aversion to a set having discrepant orderings is a special case of an aversion to a set having more than one organization of a given kind. Certain other possible explanations appeared to be ruled out by the data.

Finally, a discussion was given of how schemata might operate to produce the above and related phenomena.

REFERENCES

ADAMS, S. Status congruency as a variable in small group performance. *Soc. Forces*, 1953, **32**, 16–22.

ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.

BENOIT-SMULLYAN, E. Status, status types, and status interrelations. *Amer. sociol. Rev.*, 1944, **9**, 151–161.

BERKOWITZ, L. Sharing leadership in small, decision-making groups. *J. abnorm. soc. Psychol.*, 1953, **48**, 231–238.

DE SOTO, C. B. Learning a social structure. *J. abnorm. soc. Psychol.*, 1960, **60**, 417–421.

DE SOTO, C. B., & KUETHE, J. L. Perception of mathematical properties of interpersonal relations. *Percept. mot. Skills*, 1958, **8**, 279–286.

DE SOTO, C. B., & KUETHE, J. L. Subjective probabilities of interpersonal relationships. *J. abnorm. soc. Psychol.*, 1959, **59**, 290–294.

DE SOTO, C. B., KUETHE, J. L., & BOSLEY, J. J. A redefinition of social desirability. *J. abnorm. soc. Psychol.*, 1959, **58**, 273–275.

EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.

EXLINE, R. V., & ZILLER, R. C. Status congruency and interpersonal conflict in decision-making groups. *Hum. Relat.*, 1959, **12**, 147–162.

FESTINGER, L. *A theory of cognitive dissonance*. New York: Row, Peterson, 1957.

GUILFORD, J. P. *Psychometric methods*. (2nd ed.) New York: McGraw-Hill, 1954.

HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

HOMANS, G. C. Status among clerical workers. *Hum. Organization*, 1953, **12**, 5–10.

HOROWITZ, L. M. Free recall and ordering of nonsense syllables. Unpublished doctoral dissertation, Johns Hopkins University, 1960.

JOHNSON, D. M. *The psychology of thought and judgment*. New York: Harper, 1955.

KAY, H., & POULTON, E. C. Anticipation in memorizing. *Brit. J. Psychol.*, 1951, **42**, 34–41.

LENSKI, G. E. Status crystallization: A non-vertical dimension of social status. *Amer. sociol. Rev.*, 1954, **19**, 405–413.

POSTMAN, L., ADAMS, P. A., & BOHM, A. M. Studies in incidental learning: V. Recall for order and associative clustering. *J. exp. Psychol.*, 1956, **51**, 334–342.

RIECKEN, H. W., & HOMANS, G. C. Psychological aspects of social structure. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 2. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 786–832.

ROSE, A. M. The concept of class and American sociology. *Soc. Res.*, 1958, **25**, 53–69.

SCOTT, E. L. Leadership and perceptions of organization. *Ohio State U. Stud., Bur. Bus. Res. Monogr.*, 1956, No. 82.

SHERIF, M., WHITE, B. J., & HARVEY, O. J. Status in experimentally produced groups. *Amer. J. Sociol.*, 1955, **60**, 370–379.

STEWARD, J. H. Evolution and process. In A. L. Kroeber (Ed.), *Anthropology today*. Chicago: Univer. Chicago Press, 1953. Pp. 313–326.

SYMONDS, P. M. Notes on rating. *J. appl. Psychol.*, 1925, **9**, 188–195.

THORNDIKE, E. L. A constant error in psychological ratings. *J. appl. Psychol.*, 1920, **4**, 25–29.

WHYTE, W. F. *Street corner society*. Chicago: Univer. Chicago Press, 1943.

(Early publication received December 7, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 24-27

EFFECTS OF SOCIAL MOTIVATIONAL VARIABLES ON CHOICE REACTION TIME OF SCHIZOPHRENICS

LEONARD D. GOODSTEIN
University of Iowa

WILSON H. GUERTIN[1] AND HAROLD L. BLACKBURN[2]
Veterans Administration Hospital, Knoxville, Iowa

MOTIVATIONAL factors affecting the performance of schizophrenics, especially social rewards and punishments, have become the focus of considerable research interest. This interest partially stems from the speculation of such writers as Cameron (1947) and Fromm-Reichmann (1950), who have attempted to account for schizophrenic withdrawal (and the concomitant psychological deficit) as a defensive response to criticism, failure, and threat. These speculations have received some empirical support from the studies of Rodnick and Garmezy (1957) and their students, who have reported that schizophrenics show behavioral deficits following censure in experimental tasks involving stimulus generalization (and discrimination) and concept formation. Olson (1958), however, using a simple digit symbol task, reported that schizophrenics responded with *enhanced* performance following censure (failure instructions), although they appeared to show more enhancement following praise (success instructions) than they did following censure. There was no evidence in Olson's results to suggest that censure had any detrimental effect upon schizophrenic performance.

Rosenbaum, Mackavey, and Grisell (1957) studied the effect of "ordinary social motivation" and "biological motivation" upon the simple reaction time of schizophrenics. The biological motivation condition involved the presentation of a moderate faradic current delivered directly to the reacting finger. This stimulation began with the auditory reaction signal and terminated with the performance of the required response. The authors reported that the schizophrenic deficit in reaction time (as compared to normal controls) under the social motivation condition was substantially eliminated under the biological motivation condition.

It should be noted that the Rosenbaum, Mackavey, and Grisell study used only a single type of socially motivating instruction, one apparently involving a mild urging to do well. The work of Rodnick and Garmezy (1957), Olson (1958), and others strongly suggests that any adequate test of such motivational instructions on performance must include at least three types or levels of instructions: positive, negative, and neutral or nonevaluative. This consideration seems especially important as the instruction used by Rosenbaum and his collaborators appears most similar to the nonevaluative condition typically found to be most ineffective in modifying the level of performance.

The purpose of the present study is to investigate further the effects of social motivation upon schizophrenic reaction time, particularly the differential effects of success, failure, and nonevaluative instructions. Following Olson (1958), it was hypothesized that both the success and failure instructions should enhance the reaction time performance of schizophrenics, with success instructions producing more enhancement than failure instructions. Again following Olson, it was hypothesized that the schizophrenics performance would *not* show any detrimental effects following failure (social punishment).

METHOD

Subjects. The experimental group consisted of 42 male patients bearing the uncontested diagnosis of schizophrenia. All were hospitalized in "open buildings" at the Veterans Administration Hospital, Knoxville, Iowa. None had received lobotomies, and the overwhelming majority were receiving a minimal dosage of tranquilizing drugs.

The control group data[3] previously reported by Shankweiler (1959), were gathered from a group of 42 male general medical patients at the University Hospitals and Veterans Administration Hospital, Iowa City. Any patient who was acutely ill, whose behavior raised the question of psychosis, who had a history of treatment for psychiatric disorders, or whose history sug-

[1] Now at the University of Florida.

[2] Now at the Veterans Administration Hospital, Lebanon, Pennsylvania.

[3] The authors are indebted to Donald P. Shankweiler for making these data available to them.

gested mental deficiency, epilepsy, severe head injury, or chronic alcoholism was excluded from the control group.

The mean age of the experimental group was 39 years, and the mean educational level was 10 years. The mean age of the control group was 38 years, and the mean educational level was 9 years. Neither the differences in age nor those in mean educational level were statistically significant ($t < 1$ in both cases).

All subjects were between the ages of 16 and 62 years, had normal use of both upper extremities, had sufficient vision to see the apparatus with ease, and were capable of sitting up for an hour of testing and of understanding and cooperating in the experiment.

Apparatus. The reaction time apparatus, which was identical with that used in previous studies (Benton & Blackburn, 1957; Benton, Jentsch, & Wahler, 1959; Blackburn, 1958; Shankweiler, 1959), provided for the presentation of one of two light stimuli to the subject, each presentation being preceded by a warning buzz of 2-seconds duration.

Procedure. All subjects were first given 30 trials of simple reaction time to familiarize them with the general nature of the task. Following the simple reaction time, all subjects were given two sets of 30 trials, each involving choice reaction to one of the two stimulus lights.

The *initial* set of 30 choice trials always was preceded by the same standard orienting instructions, which emphasized that only the key with the illuminated light above it should be depressed. As a check on whether or not the subject understood the instructions, 5 practice trials, which were *not* counted in the 30 trials of the initial set, were made part of these orienting instructions. Any response completed in .100 second or less, or any depression of the key on the incorrect side, regardless of whether the correct key was depressed before or after, was considered an anticipatory error and an additional trial was added at the end of the set. Only the times of the correct responses were recorded.

The *final* set of choice trials was always preceded by one of three sets of motivating instructions: nonevaluative, success, or failure. The schizophrenic and control groups were each randomly divided into three subgroups based upon which set of motivating instructions they were given. A rest period of approximately 2 minutes was permitted each subject, after which one of the three following motivating instructions was given:

Nonevaluative Instructions: We'll record another set of reaction times. We always make at least two determinations of this choice reaction time so that we have more reliable results. We get more accurate readings that way. That's why we have so many repetitions. Now let's try this choice test again. The same instructions apply as before. Let's have a few warmups first.

Success Instructions: You've done very well. Your reaction time is quite a bit faster than average. It's really a very good performance. Your average time was about a quarter of a second. For most people it's about a half of a second. That's very good. Now let's try this choice test again. Sometime during the test I'll tell you how you're doing. [After Trial 10 and Trial 20, the experimenter assured the subject that his performance was still "very good."] Let's have a few warmups first.

Failure Instructions: You haven't done very well. Your reaction time is slower than average. It's slower than it should be. Your average time was about half a second. For most people it's about a quarter of a second. So you see, yours is pretty slow. Now let's try this choice test again. Sometime during the test I'll tell you how you're doing. [After Trial 10 and Trial 20, the experimenter commented that the subject's performance remained slow.] Let's have a few warmups first.

These motivating instructions were followed by three warmup trials and then by the 30 trials of the final set.

RESULTS

The means and *SD*s of the 30 initial trials and the 30 final choice reaction times for the six subgroups are presented in Table 1. Also presented in Table 1 are the means and *SD*s of the *gain* scores, which were obtained by subtracting the mean final reaction from the mean initial reaction time for each subject. These gain scores were the criteria used in the major analysis.

As might be expected, the schizophrenic subjects were consistently slower than the corresponding controls. The three subgroups of schizophrenics and the three subgroups of controls both showed small and nonsignificant ($F < 1$ in both cases) differences in initial choice reaction time.

An inspection of the gain scores in Table 1 indicates that, under the nonevaluative instructions, both the schizophrenics and controls

TABLE 1

MEANS AND *SD*s OF REACTION TIME SCORES (IN MILLISECONDS) FOR THE SCHIZOPHRENICS AND CONTROLS

Group	Score	Instruction Subgroup		
		Nonevaluative	Success	Failure
Schizophrenic ($N = 42$)	Initial *M*	377	372	359
	SD	52	99	55
	Final *M*	376	355	324
	SD	65	87	41
	Gain *M*	1	17	35
	SD	36	31	37
Control ($N = 42$)	Initial *M*	301	310	315
	SD	44	30	49
	Final *M*	299	295	285
	SD	43	28	48
	Gain *M*	2	15	30
	SD	22	15	31

showed no marked change in reaction time. Success instructions and failure instructions both resulted in reductions in reaction time for both groups, with the failure instructions producing the larger reduction.

The gain scores were analyzed by a simple two-by-three factorial analysis of variance (Lindquist, 1953, p. 207ff.)[4] with the following results:

1. The diagnosis-by-instruction interaction was nonsignificant ($F < 1$ for 1, 2 *df*), indicating that the instructions did not *differentially* affect the gain scores of the schizophrenics and the controls.

2. The main effect of diagnosis for the instruction subgroups combined was nonsignificant ($F < 1$ for 1, 78 *df*), indicating that there were no differential gains by the schizophrenic and the control groups as a function of the instructions.

3. The main effect of instructions over the combined diagnostic groups was highly significant ($F = 6.89$ for 2, 78 *df*, $p < .01$), indicating that the instructions were significantly different in their effect upon the gain scores.

The simple effects of instructions were then analyzed by means of *t* tests, using the pooled error term from the overall analysis, with the following results:

1. The difference in mean gain between the failure and the nonevaluative instruction groups was highly significant ($t = 3.70$, $p < .01$).

2. The difference in mean gain between the success and nonevaluative instruction groups was not quite significant ($t = 1.75$, $p < .10$).

3. The difference in mean gain between the success and failure instruction groups was also not quite significant ($t = 1.93$, $p < .10$).

Since the initial reaction times of the schizophrenic subjects were significantly slower than those of the control subjects, the dependence of the gain scores on the initial reaction time was tested. Rank-difference correlations between initial and gain scores were computed for each of the six diagnostic-instruction groups. Three of the coefficients were positive and three were negative. None was significant.

[4] An application of Hartley's test for homogeneity of variance (Walker & Lev, 1953, p. 192) indicated no significant differences in homogeneity of variance.

DISCUSSION

The results of the present investigation are noteworthy in that they offer striking evidence that the performance of schizophrenics can be improved by verbally presented social reinforcers, although the absolute level of the improved performance is still defective. The often reported failure of schizophrenics to respond to social motivation is not confirmed in this investigation. As hypothesized, both the success and failure instructions resulted in an enhancement of the choice reaction time of the schizophrenic subjects as well as the control subjects, although only the result with failure was at an acceptable level of significance. Contrary to expectation, the failure or social punishment instructions resulted in *greater* enhancement of the performance of both groups than did the success condition, although this finding is *not* statistically reliable. As hypothesized, there was *no* evidence to support the contention that schizophrenics show marked detrimental effects following socially punishing instructions. On the contrary, the schizophrenic failure group showed the largest mean gain score of any of the six groups of subjects.

The present results are in striking agreement with those of Shankweiler (1959), who studied the effect of these same motivating instructions on the choice reaction time of brain damaged subjects. The mean gain scores of the three subgroups of schizophrenic subjects used in the present investigation are within five milliseconds of the means of Shankweiler's brain damaged subjects; the means for the initial and final reaction times are also remarkably similar. Benton, Jentsch, and Wahler (1959), using different subjects on the same apparatus, previously demonstrated a similarity between brain damaged and schizophrenic subjects in both simple and choice reaction time under nonevaluative instructions. It therefore appears that schizophrenic and brain damaged patients may show not only a similarity in general reaction time performance, but also a similarity of response to these motivating instructions.

Studies of success and failure in normals (e.g., Steisel & Cohen, 1951; Williams, 1955) resemble the foregoing in that they have fairly consistently reported that mild failure in relatively simple tasks facilitates performance

to a larger extent than success, and the present findings with the control subjects are rather consistent with this. The results of the studies of success and failure in schizophrenia have been far from consistent. While Rodnick and Garmezy (1957) reported facilitation with success and decrements with failure, Olson (1958) reported facilitation with both success and failure, with the former more facilitating than the latter. The present results with schizophrenics are rather like those expected with normal subjects in that the failure resulted more in facilitation than success, although the difference is not statistically reliable. Leventhal (1959), using a verbal conditioning task, also reported social punishment ("not so good") as a more effective reinforcement for schizophrenics than social reward. Since these quite discrepant results have been gleaned from studies involving, at the very least, different tasks, different instructions, different subjects, and different experimenters, the discrepancies may simply be a function of any or all of these differences. What appears necessary is a systematic program of research in schizophrenia directly attacking these apparent contradictions.

SUMMARY

The main purpose of the present study was to determine the effectiveness of three types of motivating instruction, success, failure, and nonevaluative, on the choice reaction times of 42 schizophrenic patients and to compare their performances with those of a similar group of 42 hospitalized, nonpsychiatric patients. All subjects received an initial set of 30 choice reaction times under standard instructions; this was followed by the differential motivating instructions and by a final set of 30 choice reaction times. The criterion score for each subject was the difference in mean time scores between his 30 initial and 30 final trials.

The schizophrenic subjects were consistently slower in reaction time than the control subjects. The analysis of the gain scores indicated that the instructions were differentially effective with the rank order of effectiveness being failure, success, and nonevaluative for both schizophrenics and controls. Only the difference between the failure and nonevaluative instructions was statistically reliable. There was no difference in gain scores between the schizophrenic and control subjects.

The similarities and the differences between the present findings and previously reported research were discussed in terms of differences in task, instructions, experimenters, and subjects. Additional research was suggested as necessary to resolve the several contradictions noted.

REFERENCES

BENTON, A. L., & BLACKBURN, H. L. Practice effects in reaction time tasks in brain-injured patients. *J. abnorm. soc. Psychol.*, 1957, **54**, 109–113.

BENTON, A. L., JENTSCH, R. C., & WAHLER, H. J. Simple and choice reaction time in schizophrenia. *AMA Arch. Neurol. Psychiat.*, 1959, **81**, 373–376.

BLACKBURN, H. L. Effects of motivation instructions on reaction time in cerebral disease. *J. abnorm. soc. Psychol.*, 1958, **56**, 359–366.

CAMERON, N. *The psychology of behavior disorders.* Boston: Houghton Mifflin, 1947.

FROMM-REICHMANN, FRIEDA. *Principles of intensive psychotherapy.* Chicago: Univer. Chicago Press, 1950.

LEVENTHAL, A. M. The effects of diagnostic category and reinforcer on learning without awareness. *J. abnorm. soc. Psychol.*, 1959, **59**, 162–166.

LINDQUIST, E. F. *Design and analysis of experiments.* Boston: Houghton Mifflin, 1953.

OLSON, G. W. Failure and subsequent performance of schizophrenics. *J. abnorm. soc. Psychol.*, 1958, **57**, 310–314.

RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: Univer. Nebraska Press, 1957.

ROSENBAUM, G., MACKAVEY, W. R., & GRISELL, J. L. Effects of biological and social motivation on schizophrenic reaction time. *J. abnorm. soc. Psychol.*, 1957, **54**, 364–368.

SHANKWEILER, D. P. Effects of success and failure instructions on reaction time in patients with brain damage. *J. comp. physiol. Psychol.*, 1959, **52**, 546–549.

STEISEL, I. M., & COHEN, B. D. The effects of two degrees of failure on level of aspiration and performance. *J. abnorm. soc. Psychol.*, 1951, **46**, 79–82.

WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

WILLIAMS, J. E. Mode of failure, interference tendencies, and achievement imagery. *J. abnorm. soc. Psychol.*, 1955, **51**, 573–580.

(Received October 20, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 28-34

WORD MEANINGS AND SELF-DESCRIPTIONS

JOHN C. LOEHLIN
University of Nebraska

In psychological research, the self-concept is rather commonly defined in terms of the words a person uses to describe himself. One person describes himself as honest, energetic, impulsive, and self-assertive; his self-concept differs correspondingly from that of a person who describes himself as sensible, shy, earnest, and inarticulate. Such an interpretation implies that the words have common meaning for the different persons concerned.

On the other hand, Kelly (1955) in his work on "personal constructs" has emphasized the differences in the way individuals use concepts about personality. Differences in self-descriptions, from this point of view, may represent differences in the meanings the persons are giving to the words used, rather than differences in the way they view themselves.

In interpreting self-descriptive material, then, it becomes an important question how much disagreement exists in the meanings subjects give to the words they use to describe themselves. That *some* such disagreement exists, no one seriously questions. But many psychologists seem willing to take self-descriptions in terms of traits (e.g., adjective checklists, trait rating scales, the typical free self-description) as a first approximation to self-concepts: How serious are the resulting errors likely to be?

In evaluating the importance of differences in the meanings given to words, it seems appropriate to consider the fineness of the distinctions that must be made with them. If self-descriptions differ only slightly from person to person, even a fairly small variation in the meanings assigned to words by subjects will create serious difficulties in making discriminations, whereas the same absolute amount of variation may be of little consequence if the self-descriptions differ widely. The practical question is: How large are differences in word meanings *relative to* differences in the use of the words by subjects to describe themselves? The present investigation represents an attempt to provide an answer to this question for a situation which is important because of its frequent use in psychological research, namely, college student subjects in a classroom setting.

Specifically, the question may be considered in the following terms: Following Osgood, Suci, and Tannenbaum (1957), Cronbach (1958), Hays (1958), and others, a set of concepts may be conceived of as occupying a multidimensional "semantic space," which is defined by the relationships among the concepts in such a way that the similarity in meaning of the concepts is represented by their closeness in the space. In such a space, concepts identical in meaning occupy the same point in space, concepts similar in meaning are close together, concepts opposite in meaning are maximally distant, etc. The self-concept may also be located in such a space, with its position defined in relation to the various other concepts; thus if the concepts defining the space include careful, honest, and emotional, and a person conceives of himself as very honest, moderately careful, and not at all emotional, his self-concept will be located close to the point occupied by honest, at a moderate distance from careful, and far from emotional.

In these terms, the question becomes: What happens when we compare the "semantic spaces" of different persons? How great is disagreement on concept meanings (represented by variation from person to person in the set of interconcept distances) compared to differences in the attribution of concepts to the self (represented by variation from person to person in the set of concept-self distances)? Given a suitable method for measuring the distances in a person's "semantic space," this question can be answered empirically for any particular population of persons.

METHOD

Subjects

The subjects were from a class in the introductory study of personality at the University of Nebraska. Most were in their second semester of psychology. All but a few were nonpsychology majors—mostly liberal arts students taking the course to complete a university

requirement. The data were gathered during a regular meeting of the course, as a class exercise.

Concepts

In order to insure that the words used in the experiment be representative of a variety of aspects of personality, random samples were drawn from a pool of 82 adjectives, which had previously been selected by the writer to represent Cattell's (1946) "personality sphere" (in turn based on a condensation of Allport and Odbert's [1936] list of all the trait names in an unabridged English dictionary). The 82 adjectives included two for each of 41 of the 42 bipolar dimensions in Cattell's most recent version of the personality sphere (1957, Appendix 2). The dimension omitted was "Marked Interest in Opposite Sex vs. Slight Interest in Opposite Sex," for which no satisfactory single-word equivalents were found.

Three samples of six concepts each were independently drawn from the 82-adjective pool by means of a table of random numbers; each sample was used with a different third of the subjects, to provide three replications of the experiment. The three samples of concepts are listed in Table 1.

Measurement of Distances in the Semantic Space

Since no one technique of measuring distances in a "semantic space" is entirely free from objections, three different methods were used.

1. *Direct rating.* As one measure of the concept-concept distances, direct ratings of the similarity in meaning of pairs of concepts were used. For each concept in turn of the group of six, the subject rated its similarity in meaning to each of the other five, on a seven-point graphic scale with the end points labeled "practically the same" and "just the opposite." The subject was permitted to use whatever criteria of similarity of meaning he wished. The 30 resulting ratings represented two of each distance, one in each direction; the two ratings were summed to yield estimates of the 15 concept-concept distances for each subject.

As a corresponding direct measure of concept-self distances, the subjects simply rated themselves on each of the six concepts, using seven-point graphic scales with the end points labeled "very much so" and "just the opposite." The rating was done twice, and the two ratings summed to yield estimates of the concept-self distances for each subject.

2. *Semantic Differential.* Osgood's "Semantic Differential" (Osgood et al., 1957) provides a means of measuring distances in a semantic space based on similarities in connotative meaning. Each of the six concepts, and the self, was rated on six seven-point bipolar graphic scales—two for each of Osgood's dimensions of value, activity, and potency. The scales used were active-passive, foolish-wise, weak-strong, good-bad, hard-soft, and calm-excitable; they were chosen on the basis of data cited by Jenkins, Russell, and Suci (1958), and the instructions used were adapted from theirs.

The projection of each concept and the self on each of the three main dimensions was obtained by combining appropriate pairs of ratings (reversing direction where necessary), and the 15 concept-concept and 6 concept-self distances estimated for each subject by computing the corresponding 21 Ds (where D^2 equals the sum of the squares of the differences on the three dimensions).

3. *Rating of role persons.* This technique emphasized denotative, rather than connotative, meanings, by using a set of persons to provide a frame of reference within which distances were measured. Each subject selected a set of 12 persons who had played specified roles in his life, by means of an adaptation of Kelly's "Role Title List" (1955, p. 227). Each of these persons was rated on each of the six concepts, on seven-point graphic scales with the end points labeled "very much so" and "just the opposite." The subject also rated the similarity of himself to each of the 12 persons on scales running from "very similar" to "just the opposite."

The persons were then considered as dimensions of a semantic space, and the ratings as projections on these dimensions, and Ds computed to yield estimates of the 15 concept-concept and 6 concept-self distances for each subject.[1]

TABLE 1
LISTS OF CONCEPTS USED

Form A	Form B	Form C
complaining	submissive	inflexible
dominant	solemn	dependent
imaginative	expressive	responsible
relaxed	considerate	tough
inconsiderate	inflexible	persevering
expressive	obstructive	irresponsible

Procedure

The various rating scales were assembled in a three-part questionnaire, with the parts corresponding approximately to the three methods described; within a given part, nonself ratings preceded ratings involving the self. This questionnaire was prepared in three forms, A, B, and C, each of which used a different set of concepts (see lists in Table 1) but was otherwise identical with the others.

An effort was made to standardize as far as possible the format of the rating scales and the wording of the instructions for the three parts of the questionnaire. These instructions included "work fairly rapidly, making each judgment on the basis of your first impression," and "do not hesitate to use the extreme ends of the scales, whenever these are appropriate," as well as cautions against omitting items, talking during the experiment, etc.

[1] Properly speaking, D can be interpreted as a distance in Euclidean space only if the persons are considered as orthogonal to one another. However, with any fairly well-distributed set of persons, D should provide a good indication of *relative* distances between concepts—which is all that is needed, since the distances are later ranked. Use of the Role Title List was designed to yield a reasonable variety of persons, including liked and disliked, same and opposite sex, authority figures and peers, etc.

Subjects were randomly assigned one of the three forms of the questionnaire, and completed its three parts in the order described. The data were collected within a 50-minute class period, with most subjects requiring between 30 and 40 minutes for the task.

Analysis of the Data

Seventy-one subjects participated in the experiment. For ease in handling the data, only complete questionnaires were used for analysis—in a few cases where only single items were omitted they were arbitrarily assigned a middle rating of 4. The arrangement of the questionnaire made a skipped last page a fairly common hazard—seven questionnaires were eliminated on this ground. Eight others were excluded for other deviations from instructions—mostly for failure to rate one or more persons in Part 3. Two subjects failed to finish in the allotted time. Of the 52 remaining questionnaires, which form the basis of the subsequent analysis, 19 were Form A, 17 Form B, and 16 Form C.

The 15 concept-concept distances and the 6 concept-self distances were computed for each subject for each method of measurement as described above. As computed, the distances were in units that differed for each of the three measurement methods. Furthermore, subject-to-subject differences in mean and dispersion of ratings, due to differences in interpreting the rating scale or differences in response style, were to be expected, and in fact were apparent in the raw data. To make comparisons across methods and subjects meaningful, therefore, the set of distances for each subject for each method was converted to rank-order form. The 15 concept-concept and the 6 concept-self distances were ranked independently, because it was desired to examine the stability of the two separately. This procedure also has the advantage of minimizing any lack of comparability of concept-concept and concept-self distances that might result from the somewhat different forms of rating used for the two in two of the methods, or from possible differences in response set for judgments involving or not involving the self.

For each form of the questionnaire, mean Spearman rank-order correlation coefficients ($\bar{\rho}$) were computed among the rankings of concept-concept and concept-self distances as follows:

1. *Consistency over all rankings.* To obtain an indication of overall consistency of ranking, despite variation in both subjects and methods, Kendall's coefficient of concordance (W) was computed over the 3 N sets of rankings in each group (where N is the number of subjects in the group, i.e., 19, 17, or 16).[2] This was done separately for the concept-concept and concept-self distances, and then each of the six resulting Ws was converted to $\bar{\rho}$ by the relation $\bar{\rho} = (mW-1)/(m-1)$, where m is the number of arrays over which W is computed (Kendall, 1948).

2. *Consistency within methods.* If the three measurement methods are measuring different things, but measuring them dependably, the overall consistency of rankings may be low, but consistency within methods should be considerably higher. Therefore, for each of the three methods, Ws were computed over the N subjects in each group, separately for the concept-concept and concept-self distances, and each W converted to $\bar{\rho}$ as above.

3. *Consistency within subjects.* Subjects may display individual differences, consistent over measurement methods, in the assignment of meaning to concepts ("personal constructs"), and it is ordinarily assumed that they display consistent individual differences in the "meanings" they assign to themselves (i.e., self-concepts). These should result in within-subject consistency among rankings that is substantially higher than consistency measured across subjects. Therefore, Ws were computed for each subject over the three measurement methods, separately for the concept-concept and concept-self distances, and these converted to $\bar{\rho}$'s as above.

4. *Consistency across subjects and methods.* To provide an unconfounded baseline against which to evaluate intrasubject and intramethod consistency, the $\bar{\rho}$ for all possible comparisons made across both subjects and methods was computed. This was done by multiplying the overall $\bar{\rho}$ by the number of comparisons involved, subtracting from this the within-person and within-method $\bar{\rho}$'s, each multiplied by the number of comparisons on which it was based, and dividing the remainder by the number of across-method-and-subject comparisons. Thus, for example, in Group A there are 1,596 possible paired comparisons among the 57 sets of rankings, with $19 \times 3 = 57$ of these within subjects, $3 \times 171 = 513$ of these within methods, and the remaining 1,026 across both subjects and methods. The computation is $(.390 \times 1{,}596 - .443 \times 57 - .449 \times 513)/1{,}026 = .358$ for the $\bar{\rho}$ for comparisons across both subjects and methods.

RESULTS

Table 2 contains the mean rank-order correlation coefficients among rankings within the categories specified.

Let us begin by examining the self-descriptions. First, do individual subjects describe themselves with any consistency whatsoever? The affirmative answer is in the left-hand part of the first row of Table 2. When $\bar{\rho}$'s are computed for individual subjects over the three measurement methods, they tend to be substantially positive: the means of these $\bar{\rho}$'s in the three subgroups are .48, .40, and .50, respectively.

The second, and more crucial question for our purposes is: To what extent do the in-

[2] No correction was made for tied ranks. The effect of such a correction would be to raise W very slightly, and to about the same extent for concept-concept and concept-self distances. In making comparisons among the methods, however, allowance may be made for the fact that ties were somewhat more frequent in the "direct rating" method than in the other two, and hence its consistency is probably slightly underestimated.

TABLE 2
CONSISTENCY AMONG RANKINGS OF CONCEPT-CONCEPT AND CONCEPT-SELF DISTANCES

Comparisons	Mean ρ Among Rankings					
	Of 6 Concept-Self Distances			Of 15 Concept-Concept Distances		
	Form A	Form B	Form C	Form A	Form B	Form C
Among all subjects[a]						
Within subjects, across 3 methods	.48	.40	.30	.44	.43	.60
Across subjects and methods	.38	.21	.40	.36	.28	.34
Among consistent subjects only [b]						
Within subjects, across 3 methods	.70	.70	.74	.58	.61	.64
Across subjects and methods	.64	.58	.62	.50	.46	.33
Within methods, across N subjects[a]						
Direct rating	.29	.34	.55	.47	.42	.38
Semantic Dif.	.30	.13	.45	.30	.31	.32
Role persons	.54	.30	.25	.57	.38	.36
3 methods combined	.37	.26	.42	.45	.37	.35
Across subjects and methods	.38	.21	.40	.36	.28	.34
Over all 3 N rankings	.38	.23	.41	.39	.31	.35

[a] N = 19, 17, and 16 for Forms A, B, and C, respectively.
[b] N = 7, 3, and 8 for Forms A, B, and C, respectively. Consistency as defined in text.

dividual subjects describe themselves *distinctively*? The usefulness of the self-concept to the personality psychologist usually presupposes that differences from individual to individual can be dependably ascertained. The relevant information in Table 2 lies in the relationship between Rows 1 and 2 on the left-hand side. The difference between consistency across methods within persons and consistency across methods across persons represents the effect of consistent individual differences. This effect is not impressively large, but it is present in all three groups.[3] Thus these subjects do appear to differ from person to person in how they describe themselves, although there is also a substantial consensus in their self-descriptions, which appears upon inspection of the data to be based on a tendency noted by Edwards (1957) and others for subjects to apply socially desirable words to themselves, and avoid socially undesirable words.

Person-to-person differences in self-description might arise, however, simply from person-to-person differences in the meanings attributed to the concepts used in self-description. Are such differences present in the data, and if so, how large are they? The answer may be found by comparing Rows 1 and 2 on the right-hand side of Table 2. Here again Row 1 reflects consistency across methods within subjects, and Row 2, consistency across methods across subjects. The difference between the two reflects consistent individual differences in the meanings assigned to concepts. And this difference is roughly of the same magnitude as that on the left-hand side of the table:[4] in other words, the concepts appear to differ in meaning from person to person fully as much as their use in self-description does.

Are these differences related? In order to shed some tentative light on this question, one concept in each group was selected (the one that appeared to show the greatest variability in concept-self distance from person to person) and the relation between its distance from the self and its distance from the other concepts in its set was evaluated by computing ρ's between them over the subjects in the group.

In general, the relationships observed suggest that a concept was more likely to be used in self-description when it was given a more favorable meaning. Take, for example, the concept *solemn* in Group B. Individuals differed considerably as to whether they located it near or far from their self-concept—for some it was one of the more self-descriptive of the concepts in its set; for others it was the least. And the variation in the use of the concept in self-description appeared to be accompanied by shifts in its meaning: those persons who judged solemn to be self-descriptive located it relatively close in the semantic space to expressive and considerate; those who did not, placed it distant from these concepts, and closer to submissive or to obstructive. In the

[3] Individual subjects having $\bar{\rho}$'s above the $\bar{\rho}$ across persons in their group number 12 out of 19, 11 out of 17, and 10 out of 16, in the three groups. The two-tailed probability (based on the binomial) of obtaining such a result by chance is .014.

[4] Its significance is greater—well beyond the .001 level—due to less variability among individuals in the consistency of these rankings.

same way, in Group A, when relaxed was seen as relatively close in meaning to expressive and imaginative, it was more likely to be considered self-descriptive than when it was far from these concepts. Or in Group C: whether dependent was considered self-descriptive or not went with its nearness to persevering and responsible, on the one hand, or to irresponsible, on the other.[5] It seems plausible, therefore, to attribute the differences in self-description at least in part to the differences in meaning that subjects attributed to the concepts.

One possible complication in the interpretation of the results was suggested by the fact that for several subjects consistency across methods was around zero, while for others it was quite high. It seemed desirable to check the general conclusions of the study among those subjects for whom the measurement appeared to be best: that is, those subjects who showed high consistency across methods, both in the measurement of concept-self distances and in the measurement of concept-concept distances. A $\bar{\rho}$ of .50 was chosen as the arbitrary cut-off point for including subjects in the "consistent" group. This was approximately at the median of each distribution. About one-third of the total sample exceeded this level of consistency on both distributions, and were included in the group of consistent subjects. The third and fourth rows of Table 2 contain $\bar{\rho}$'s based on this group. It will be seen that despite the now considerably higher $\bar{\rho}$'s (particularly for the concept-self distances), the differences between within- and across-person $\bar{\rho}$'s remain of roughly the same order of magnitude on the two sides of the table. Again, it appears that the shift in concept meanings from person to person is about the same as the change in self-descriptive use of the concepts.

Finally, it is of some interest to examine the consistencies within the three measurement methods. The relevant $\bar{\rho}$'s are in the lower part of Table 2. There appears to be some tendency for the "direct rating" and "role persons" techniques to yield higher within-method than across-method $\bar{\rho}$'s, but there is an exception among the concept-self comparisons in each case. The Semantic Differential may produce slightly less agreement than the other techniques, but again, this is not uniformly true in all the groups. In general it appears that the three methods agree reasonably well on what they are measuring, in that within-method $\bar{\rho}$'s are not, on the average, greatly in excess of across-method $\bar{\rho}$'s. (This is particularly true for the concept-self distances.)

[5] The relationships mentioned are based on ρ's ranging from .45 to .79.

DISCUSSION

If it is true, as the present results appear to indicate, that consistent individual differences in self-description are on the whole no larger than consistent individual differences in the meanings of the words used in self-description, then obviously inferences about the self-concept from such self-descriptions are very hazardous.

Some restrictions on the generality of these results should be noted at the outset.

First, the subjects were college students in a psychology class. However, one might expect that in a more heterogeneous population consensus on word meanings and on self-concepts would both decrease somewhat, leaving the nature of the problem of interpreting self-descriptions much the same.

Second, the results apply to self-descriptions in terms of adjectival trait names. Presumably if the words are defined for the subject, as is done in some rating scales, the variability in meaning from person to person should decrease (to an extent limited by the fact that the words used in definition are likely to be themselves ambiguous). Reduction in variability of meaning may also result from the incorporation of the adjectives in sentences, as is done in many *Q* sorts and inventories, but this need not always be true. In "I am more sensitive than most other people," the word *sensitive* is probably just as variable in meaning as when it is used alone. However, the findings of the present study apply most directly to self-description by such means as adjective checklists or simple trait rating scales.[6]

[6] It should be made clear that these results do not imply that adjective checklists and similar self-rating devices are without value. For some uses of such instruments, the fact that individuals may differ in the meanings they assign to words does not present serious difficulties. For example, in strictly empirical prediction problems it may be immaterial whether the instrument works because it measures self-concepts, personal constructs, or some mixture of the two. In

Third, the results apply to self-description on a single occasion—the typical research situation; implications are necessarily indirect for, say, a clinical setting in which self-concepts or personal constructs emerge gradually over a series of interviews.

Finally, the results depend on the validity of measuring meanings in terms of distances in a semantic space (the fairly good agreement across the three measurement methods suggests that the results should be independent of any *particular* method of measuring such distances). The merits of a model of this kind have been discussed by Osgood et al. (1957) and others. A particular advantage for the present study is that it permits evaluation of consistency of meaning of self and other concepts *in the same terms*, i.e., as the stability across persons or methods of the rank order of a set of distances in a given semantic space.

This last is important. "Everyone knows" that people's agreement on the meaning of words is greater than their agreement on their self-descriptions (contrary to the present findings), but this is agreement in terms of *dictionary definitions*, that is, in terms of distances to that set of concepts most stably related to the concept in question, not in terms of position in the semantic space as a whole. A relevant example can be drawn from the data of this study. It so happened that the two adjectives responsible and irresponsible were included in the sample of six concepts in Group C. The data suggest that the subjects were in high agreement as to the relationship of these concepts *to each other*. However, the subjects still differed substantially on how the pair related to other concepts in the set. No doubt if dependable and untrustworthy had been included along with irresponsible and responsible, one would have found highly consistent relations among the four concepts and perhaps have concluded that the subjects "agreed on what they mean"—but do they? The writer would insist that it is the general relationships in the semantic space—the relationship between, say, independent and responsible, rather than that between irresponsible and responsible—which constitute the "meaning" with which the psychologist is usually concerned. And it is *this* kind of meaning which, in the present experiment, differs from person to person as much as the use of the concepts in self-description does.

What is the process which produces these related differences in self-description and concept meaning? If (to recur to an earlier example) solemn means something like expressive to the person who uses it to describe himself, and something like obstructive to the person who does not, how does this come about? One line of interpretation is in terms of the theory of self-consistency proposed by Lecky (1945), and adopted in modified form by Murphy (1947), Rogers (1959), and others. These writers emphasize the fact that evidence about the self received from external sources tends to be modified to fit the person's existing concept of himself. In the present case, such an explanation might take either of two forms. It might be that a person first interprets solemn in a favorable light, as meaning something like expressive, and thereafter is more willing to apply it to himself than is someone who interprets it as more like obstructive. Or it might be that he first accepts solemn as applying to himself (because people have called him that) and then tends to give the word a particular interpretation in line with his general view of himself. The present data do not offer a basis for decision between these alternatives. However, in either case the general conclusion is the same: we know little about what it means when a person calls himself *solemn* unless we know what he means by the word—and one basic implication of the present study is that it isn't at all safe to assume that he means the same thing by it that we do.

SUMMARY

College students made a series of ratings designed to determine the similarities of meaning among a set of six adjectival concepts, as well as the relation of the subject's self-concept to each of the concepts in the set. Three different techniques of measuring meaning were used. Inter- and intrasubject consistency in word meanings and self-descriptions

other cases, the assumption of consistency within the individual may be sufficient: one may be able to make meaningful comparisons of overall discrepancies between, say, self-concept and self-ideal, even if the dimensions along which such discrepancies are measured vary from person to person.

were examined. For three groups of subjects using different sets of adjectives, consistent individual differences in the meanings attributed to words proved to be of the same order of magnitude as consistent individual differences in the use of these words in self-description. It was suggested that this implies considerable hazards in using self-descriptions of this nature to define the self-concept.

REFERENCES

ALLPORT, G. W., & ODBERT, H. S. Trait-names: A psycho-lexical study. *Psychol. Monogr.*, 1936, **47** (1, Whole No. 211).

CATTELL, R. B. *The description and measurement of personality*. Yonkers-on-Hudson, N. Y.: World Book, 1946.

CATTELL, R. B. *Personality and motivation structure and measurement*. Yonkers-on-Hudson, N. Y.: World Book, 1957.

CRONBACH, L. J. Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford, Calif.: Stanford Univer. Press, 1958. Pp. 353–379.

EDWARDS, A. E. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.

HAYS, W. L. An approach to the study of trait implication and trait similarity. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford, Calif.: Stanford Univer. Press, 1958. Pp. 289–299.

JENKINS, J. J., RUSSELL, W. A., & SUCI, G. J. An atlas of semantic profiles for 360 words. *Amer. J. Psychol.*, 1958, **71**, 688–699.

KELLY, G. A. *The psychology of personal constructs*. Vol. 1. *A theory of personality*. New York: Norton, 1955.

KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.

LECKY, P. *Self-consistency: A theory of personality*. New York: Island Press, 1945.

MURPHY, G. *Personality: A biosocial approach to origins and structure*. New York: Harper, 1947.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana, Ill.: Univer. Illinois Press, 1957.

ROGERS, C. R. A theory of therapy, personality, and interpersonal relationships, as developed in the client-centered framework. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 3. *Formulations of the person and the social context*. New York: McGraw-Hill, 1959.

(Received October 26, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 35–40

THE EFFECTIVENESS OF UNANTICIPATED PERSUASIVE COMMUNICATIONS

JANE ALLYN AND LEON FESTINGER[1]
Stanford University

It seems plausible to suppose that an attempt to persuade a person to change his opinion on some issue would be more effective if the persuasive communication were unexpected than if the person anticipated the influence attempt. At least many people find this to be plausible. Lazarsfeld, Berelson, and Gaudet, for example, state "If we read or tune in a speech, we usually do so purposefully, and in doing so we have a definite mental set that tinges our receptiveness . . . . This mental set is armor against influence. The extent to which people, and particularly those with strong partisan views, listen to speakers and read articles with which they agree in advance is evidence on this point" (1948, p. 152), and Festinger says: "It seems clear that the avoidance and evasion of material which might produce or increase dissonance depends on anticipations (probably unverbalized ones) about the material or on preliminary assessments of the material" (1957, p. 158).

There are, however, virtually no data either to challenge or to support such statements. A study by Ewing (1942) comes closest to being pertinent. Ewing exposed two groups of subjects to a persuasive communication that supported a very extreme point of view. The only difference in treatment between the two groups was in the introduction to the persuasive communication. This introduction led one group to expect that the communication would support the extreme view which it actually espoused. For the other group, the introduction led the subjects to expect that the communication would agree with their existing views. The results indicated that those subjects who expected the communication to disagree with their opinions were actually less influenced by it. In other words, the persuasive communication was more effective if the audience falsely anticipated that it would support their existing views. This study does not deal with the effectiveness of a communication to an entirely unprepared audience, but it does point to the operation of some anticipatory resistance by an audience which expects to hear a disagreeing speech.

One may, of course, raise the further question as to the exact nature of the process by means of which the prepared audience is able to resist the effects of the attempted persuasion.

Some understanding of this process may be obtained from Festinger's theory of cognitive dissonance (1957). We may say that a persuasive communication which argues for an opinion different from that held by the audience creates dissonance in the listeners. Specifically, dissonance is created between the opinion the listener holds and his knowledge of the arguments favoring a contrary view. Since a person who reads a persuasive communication, or is a listener in an audience, cannot attempt to influence the source of the communication, there are only two immediate ways in which he can reduce this dissonance. He can change his opinion to a position closer to that advocated by the communication or he can reject and derogate the communication and the communicator.

What, then, are the effects of being prepared or unprepared to hear a persuasive communication? If the conditions of exposure to a communication are such as to guarantee attentiveness, it seems plausible to assume that, hearing the same speech, the same amount of dissonance is introduced whether or not the person is prepared for it. In such a case, the effect of preparedness, if any, would be on the particular mechanism used to reduce the dissonance, namely, whether the person more readily changes his opinion or rather rejects the source of communication.

There is no rigorous derivation to be made

[1] The authors would like to express their thanks to Nathan Maccoby for his help in writing the persuasive communication and in the conduct of the experimental sessions. We are also indebted to John G. Caffrey and Curt R. Demele for their help and cooperation in obtaining subjects.

here but one may argue that, if a person anticipates hearing a communication that will disagree with an opinion he holds strongly, he will approach the situation with hesitancy, suspicion, and perhaps some hostility. If he does approach the situation in this way, then it seems natural to expect that his first and easiest reaction will be to reject the communicator.

Thus, being prepared for the communication would not, in a sense, make the communication less effective. It would simply alter the way in which the dissonance is reduced. Those prepared for the communication would tend less to change their opinion and tend more to reject the communicator. The study by Ewing (1942) gives some support to this analysis of the situation. He found that the discrepancy between the subjects' initial opinion and the position they expected the communication to espouse was the major determinant of how biased they felt the communication to be. In other words, irrespective of what the communication actually said, if they expected disagreement, they later felt the communication to be more biased. Anticipation of disagreement seems to have led to more rejection of the communication.

The present experiment was designed to throw more light on this process by comparing two groups exposed to the same persuasive communication, one group expecting a disagreeing persuasive communication and the other group not expecting a persuasive communication at all.

PROCEDURE

A questionnaire was given to 128 students of the Palo Alto High School in order to discover a topic that was important to these students and on which their opinions were relatively extreme and homogeneous. This questionnaire was administered in the classrooms by the teachers who described it as a "Youth Survey." The subjects were assured that their replies would be confidential. The questionnaire included questions about the control of teenage drivers, teenage curfew regulations, and the treatment of communists in this country. All questions had four alternative answers: "agree strongly, agree slightly, disagree slightly, disagree strongly." The subjects were also asked to rank the three topics according to how important they were. On the basis of these rankings of importance and the clustering of their opinions toward one extreme, the issue of controlling teenage automobile driving was chosen for use in the actual experiment. The four questions concerning this topic on this questionnaire also served as our first measure of opinion.

Only 91 of these subjects were present at the experimental session 2 weeks later. The data from four of these were not analyzed; two arrived too late, and two were discarded because their responses were mainly attempts at humor. Of the remaining 87 subjects, 53 were females and 34 were males. The mean age was 15.7 years, with a range from 14–18.

The experimental session was conducted in the high school auditorium. As each student entered, he was given a questionnaire booklet. The students were asked to sit apart from each other and not to open the booklets until instructed to do so. The experimenter then explained to them that they were participating in an experiment conducted by the Stanford Psychology Department and assured them that their answers would be confidential and would have no effect on the grades they would receive in their classes. The subjects were then asked to open their booklets to the first page and read the printed instructions. There were two different sets of instructions and the booklets had been randomized in this regard when they were passed out. In one condition, "Opinion Orientation," subjects read the following:

> We are very much interested in studying the opinions and attitudes of high school students on the teenage driving problem. We particularly want to get your reactions to a talk by an expert on the subject. We have asked Mr. Nathan Maccoby, who is very well known for his reports and articles on this problem, to come here today and talk to you. In his articles, Mr. Maccoby has stated his very strong opinion that teenagers are a menace on the roads and should be strictly controlled by effective new laws. It is important that you pay close attention to him and to what he has to say, as you will be asked to give your opinions on the problem after he has finished his talk.

In the other condition, "Personality Orientation," subjects read the following:

> We are very much interested in studying how members of an audience form impressions of the personality of a speaker. We particularly want to see how high school students will "size up" the personality of an adult who is speaking to them on a serious subject. We have asked Mr. Nathan Maccoby, who is considered to be an expert in his field, to come here today and talk to you. We have asked him to speak on the topic that he knows the best, so as to put him more at his ease. It is important that you pay close attention to him and to what he has to say, as you will be asked to give your impressions of his personality after he has finished his talk.

When they had finished reading the instructions, the speaker was introduced. He delivered a speech that stressed the teenager's lack of a mature sense of responsibility and advocated stricter legal measures for the prevention and control of teenage driving. The communication was more extreme in the direction of control than the opinions of any of the students. After the speech, the experimenter thanked the speaker and instructed the students to answer the questions in their

booklets. The booklets contained the same four questions on control of teenage driving which had been used previously together with other questions designed to measure rejection of the speaker, importance of the topic, and other items. These will be discussed in more detail under "Results." At the conclusion of the session the experiment was explained in simplified terms and any questions the students had were answered. The session took approximately 45 minutes.

RESULTS

Effectiveness of the Experimental Manipulation

The purpose of one of the sets of instructions was to create an orientation toward the personality of the speaker so that the actual persuasive communication would be unanticipated. The other set of instructions was intended to create an orientation toward the opinion of the speaker together with the expectation that the speaker would disagree with their own opinion. To check on whether or not the two different orientations were successfully created, the following question was asked:

> The psychologists who are doing this experiment are mainly interested in
> (A) what kinds of opinions people have
> (B) how people judge other people's personalities.

If our manipulation was successful, we would expect the majority of Opinion Orientation subjects to check the first alternative and the majority of Personality Orientation subjects to check the second. Out of a total of 41 subjects who received the Opinion Orientation instructions, 80% said the experiment was concerned with opinions. Out of the total of 46 subjects who received the Personality Orientation instructions, 63% said the experiment was concerned with judgment of personality. The difference between the two conditions is, of course, highly significant statistically. Thus it appears that our experimental manipulation was effective in creating different orientations toward the speaker in the two groups.[2]

It should be noted in the instructions to the subjects that in each of the two conditions an effort was made to introduce the speaker as equally expert. This was done because we wanted to keep constant the magnitude of dissonance occasioned by the persuasive communication. The following question was asked to check on this point:

> When we introduced Mr. Maccoby, we said that he is considered to be an expert in his field. How much do you think he really knows about the problem he discussed?
> (A) He is a real expert, who is very familiar with the problem.
> (B) He knows quite a lot about the problem.
> (C) He knows about as much as most people do about the problem.
> (D) He knows less than most people about it.
> (E) He knows very little, and a lot of his information is wrong.

The results show that we were successful in keeping the expertness of the speaker constant between the two conditions. In the Opinion Orientation condition 78% said the speaker was a real expert or knew quite a lot about the problem. In the Personality Orientation condition the comparable figure is 80%. Clearly, there are no appreciable differences between the two conditions.

Since the purpose of the experimental manipulation was to affect the orientation of the subjects but to keep constant the magnitude of dissonance, it is also important to know whether or not the subjects in the two conditions paid equal attention to what the speaker said. For this purpose a number of factual recall questions were asked. The results show no difference between the conditions in amount of material recalled. Since the two conditions were the same with respect to both attention paid to what the speaker said and degree of expertness attributed to him, we may regard the experimental manipulation as having created a difference in orientation or preparedness on the part of the subjects while holding constant the magnitude of dissonance introduced by the persuasive communication.

Comparison of the Two Experimental Conditions

We expected subjects in the Opinion Orientation condition to reduce dissonance mainly by rejecting the communicator and expected those in the Personality Orientation condition to reduce dissonance mainly by changing their opinion in the direction advocated by the communication. Change scores were calculated as follows: If a subject changed his answer to an adjacent category, e.g., from "agree strongly" to "agree slightly,"

[2] There were no significant differences in the responses given by boys and girls on any of the measures in the study. Therefore, the data were analyzed for the group as a whole.

he received a score of 1; if he changed to an answer two steps away, e.g., from "disagree strongly" to "agree slightly," he received a score of 2; since we used a four-point scale, the maximum possible change was 3. Changes in the direction of the position advocated by the communication were given a plus sign, and the changes were summed algebraically for the four questions to obtain a single score. Thus, if a subject changed from "disagree strongly" to "agree slightly" on one question in the direction advocated by the communication (a change of +2); changed from "agree strongly" to "agree slightly" on another question in the direction opposite to that advocated by the communication (a change of −1); and did not change at all on the other two questions, his opinion change score would be +1.

Rejection of the communicator was measured by the following question:

> Do you think Mr. Maccoby covered all the facts pretty well, or do you think he was unfair or biased in some way?
>
> (A) He was very unfair and biased.
> (B) He was somewhat unfair; he didn't give all the facts.
> (C) He was very fair; he covered all the facts pretty well.

Table 1 presents the data for opinion change and rejection. Two indices of opinion change are presented in the table: the average change and the percentage of subjects changing appreciably (two or more points) in the direction advocated by the communication. Both of these measures show differences between the two experimental conditions in the predicted direction. Subjects in the Personality Orientation condition change their opinion more in the direction advocated by the communication than do subjects in the Opinion Orientation condition. While the difference in average change of opinion is not at all significant, the percentage who change appreciably is significant statistically (chi square = 5.71, $p = .02$).[3] The difference between the two experimental conditions on rejection of the communicator is also in the expected direction. Subjects in the Opinion Orientation condition are more likely to reject the communicator. The difference is statistically significant (chi square = 4.06, $p = .05$).

TABLE 1

OPINION CHANGE AND REJECTION OF COMMUNICATOR FOR THE TWO EXPERIMENTAL CONDITIONS

	Experimental Condition	
	Opinion Orientation ($N = 41$)	Personality Orientation ($N = 46$)
Average change of opinion	+.49	+.63
Percentage changing appreciably[a]	20%	43%
Percentage saying communication was very or somewhat biased	80%	61%

[a] An appreciable change is defined as a change of two or more points in the direction of the communication.

In other words, our expectations were borne out. If the subjects are led to expect that a persuasive communication will disagree with them, they reject the communicator more and are less influenced than subjects who are, so to speak, caught unprepared. The obtained differences, however, while significant statistically, are rather small. Let us, therefore, examine the data further to see what additional light can be shed on the process with which we are dealing.

Initial Opinion and Importance of Issue

The topic of teenage driving was chosen for the experiment because the initial opinions of the high school students were more extreme on this issue than on either of the other issues that we pretested. Theoretically, of course, we would expect that the differences between our two experimental conditions would be greater for people who held opinions very divergent from the position advocated by the communication since more dissonance is introduced for them. Although opinions on the issue of teenage driving were relatively extreme, still, only a minority of the subjects held entirely one-sided opinions on the issue.

The four questions used as the measure of opinions about teenage driving on both questionnaires were as follows:

1. Since most accidents involving teenagers happen at night, people between the ages of 16 and 20 should only be allowed to drive during the daytime.
2. Tests for obtaining a license to drive should be

[3] All p values reported are two-tailed.

made much harder for teenagers and should also test for judgment and knowledge of the traffic laws.

3. I think that the minimum age for obtaining a driver's license in California should be raised to 18. [Note: present age is 16.]

4. If a person between 16 and 20 years of age receives a ticket for speeding or reckless driving, his license should be revoked until he is 21.

If we define an extreme opinion as disagreement (either strong or slight) on all four questions, we find that only 16 of the 41 subjects in the Opinion Orientation condition and 16 of the 46 subjects in the Personality Orientation condition initially held such extreme opinions.

Another variable that should affect the results of the experiment is the importance of the issue to the subject. The more important the issue of teenage driving is to a particular person, the greater should be the magnitude of dissonance created in him by the persuasive communication. In order to obtain a measure of the importance of the issue to each subject, we asked the following question:

How important is it to you to be able to have a license to drive a car?
(A) very important
(B) quite important
(C) not too important
(D) not at all important

Ideally we would have wanted to analyze the data from the experiment separately for the effect of extremeness of opinion and the effect of importance of the issue. Unfortunately, it turns out that there is a very appreciable correlation between these two variables. In both experimental conditions equally, those who hold extreme opinions also overwhelmingly state that the issue is important to them. Hence, we cannot look at the effect of each of these variables separately Consequently the data were analyzed for those with extreme opinions versus those with moderate opinions on the issue, ignoring the specific responses on importance. It should be remembered, however, that these two go hand-in-hand in this sample. Since, theoretically, both of these variables should affect the data in the same manner, the interpretation will be, of necessity, slightly ambiguous. That is, the greater the difference between initial opinion and the position advocated by the communication, and the greater the importance of the issue, the greater will be the magnitude of dissonance introduced by the communication. Hence, we cannot separate the effects of these two variables in our data.

Let us, then, examine the data to see the effect of the combined action of both variables. Table 2 presents the data on opinion change separately for those who initially held extreme opinions and those who initially held moderate opinions for each experimental condition. It should be remembered that a subject was labeled as holding a moderate opinion if he initially agreed with the position advocated by the communication on one or more of the questions.

It is clear from Table 2 that the difference between initial opinion and position advocated by the communication is a major variable. For those who held extreme opinions initially (and for whom the issue was more important) there is an average change of +2.31 in the Personality Orientation condition as compared to an average change of only +.81 in the Opinion Orientation condition. The difference on the measure of percentage who change appreciably is also similarly large. Both differences are significant at better than the 2% level. For those whose initial opinion was moderate (and the issue was less important) there is practically no difference between conditions and, indeed, practically no impact of the communication at all. The average change for the Opinion Orientation group is only +.28 and the average for the Personality Orientation group is even slightly in the opposite direction from that advocated by the communication. In other words, the entire difference that we found between

TABLE 2
OPINION CHANGE IN RELATION TO INITIAL OPINION

	Experimental Condition	
	Opinion Orientation	Personality Orientation
Extreme Initial Opinion		
Average Change	+.81	+2.31
Percentage changing appreciably	19% (N = 16)	60% (N = 16)
Moderate Initial Opinion		
Average Change	+.28	−.27
Percentage changing appreciably	20% (N = 25)	30% (N = 30)

the experimental conditions is attributable to those subjects for whom the persuasive communication introduced considerable dissonance.

We would also expect the difference between the two experimental conditions on the measure of rejection to be greater for those with initially extreme opinions. Surprisingly, however, this does not turn out to be the case. There are simply no differences between those holding extreme opinions and those holding moderate opinions in the percentage who feel the communicator was biased. In the Personality Orientation condition these figures are 62.5% and 60%, respectively, and the comparable figures for the Opinion Orientation condition are 81% and 80%. There is, of course, no reason to expect a difference in the Personality Orientation condition where dissonance was reduced mainly by changing opinion. There is reason, however, to have expected a difference in the Opinion Orientation condition. It is conceivable, however, that the overall level of rejection was already so high in this condition (80%) that a virtual ceiling had been reached as far as our relatively crude measure is concerned. This cannot, of course, be settled with the present data. The only thing that can be said is that, if there is not a ceiling phenomenon here, there is a theoretical difficulty involved in interpreting the rejection data.

SUMMARY

Eighty-seven high school students who were in favor of allowing teenagers to drive with few restrictions were presented with a communication advocating strict control of young drivers. One group was given an orientation to attend to the speaker's opinions and was informed of his topic and his point of view in advance; the other group was given an orientation to evaluate the speaker's personality and was not told the topic of the speech or the speaker's point of view. It was found that subjects who were forewarned of the nature of the communication changed their opinions less and rejected the communicator as biased to a greater degree than unprepared subjects. Differences in amount of opinion change between prepared and unprepared subjects were greater among those holding extreme opinions initially.

REFERENCES

EWING, T. A study of certain factors involved in changes of opinion. *J. soc. Psychol.*, 1942, **16**, 63–88.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston, Illinois: Row, Peterson, 1957.

LAZARSFELD, P., BERELSON, B., & GAUDET, HAZEL. *The people's choice.* (2nd ed.) New York: Columbia Univer. Press, 1948.

(Received November 12, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 41–43

PERSONALITY AND VERBAL CONDITIONING EFFECTS[1]

GEORGIA BABLADELIS[2]
University of Colorado

As a body of evidence about verbal conditioning has accumulated, investigators have begun to discuss its implications for understanding the psychotherapeutic process. An important question that arises in connection with psychotherapy is that of what happens in the psychotherapeutic process to produce change in the behavior of the client.

Increased attention has been given to the verbal behavior of the client in attempts to solve both practical and theoretical problems here (Mowrer, 1953; Raimy, 1948; Rogers & Dymond, 1954; Shaffer, 1947; Shoben, 1949; Wolpe, 1958). Conceivably, verbal behavior may serve as either cause or effect of both affective and instrumental responses. If this is true, it is logical to expect that the manipulation of verbal patterns as mediators is one source of change in the range of both affective and instrumental response patterns of the client, and thus of the larger patternings of his interpersonal relationships.

In verbal conditioning studies, the experimenter attempts to manipulate verbal mediators. Krasner (1958) has presented a thorough review of the studies in this area, pointing out that this group of studies follows Skinner's paradigm in that "the dependent variables are the subject's verbal behavior, and the independent variables are generalized conditioned reinforcers, intended to bring verbal behavior under the control of the examiner."

Certain deficiencies exist in the general design of verbal conditioning studies reported, especially when they are considered as analogues to psychotherapy. Usually, there is only a strained or distant relationship between the experimental setting (noninterview) and the therapeutic setting (interview), so the generalization of results is limited. Similarly, little attention is ordinarily given to personality variables, and when attention is given to them, complicated problems arise. Those experimenters (Sarason, 1958; Taffel, 1955) who have studied the relationship of personality variables to verbal conditioning effects have not employed an interview situation, nor have they selected a response class which appears to be meaningful to the subjects. Those experimenters (Salzinger & Pisoni, 1958; Verplanck, 1955) who have employed the interview or who have attempted to condition a response class which seems to be a more integral part of the subject's verbal behavior have not studied the relationship of personality variables to verbal conditioning effects.

The present study is an attempt to overcome these restrictions by using interviews with the subjects who vary along the measured personality dimension of Autonomy. It is an attempt to employ successfully the interview situation, to condition a response class that appears to be a meaningful aspect of subject's verbal behavior (i.e., self-descriptive statements), and to demonstrate that a personality variable—Autonomy—imposes a limitation on verbal conditioning effects. The basic hypothesis is that the relationship between scores on Autonomy and verbal conditioning effects is linear and negative.

Interaction and main effects were tested in an analysis of variance design.

METHOD

Edwards (1954) has given the following definition of Autonomy:

> To be able to come and go as desired, to say what one thinks about things, to be independent of others in making decisions, to feel free to do what one wants, to do things that are unconventional, to avoid situations where one is expected to conform, to do things without regard to what others may think, to criticize those in positions of authority, to avoid responsibilities and obligations (p. 14).

The Autonomy scale of the Edwards Personal Preference Schedule is so constructed that a high scorer is one who is expected to manifest many of the characteristics

[1] This article is based on a dissertation submitted to the University of Colorado in partial fulfillment of the requirements for the PhD.

[2] The author is indebted to Victor Raimy, Milton Lipetz, and Donald Stilson for their helpful suggestions and encouragement throughout the study.

Now at the Guidance Clinic, Alameda County Probation Department, Oakland, California.

described and a low scorer is one who is expected to manifest few or none of them. The EPPS was administered to 102 students in undergraduate psychology courses at the University of Colorado. The Autonomy scale was arbitrarily divided into four segments by cutting the obtained distribution at the mean and at one standard deviation on either side. The tests of all 102 subjects were sorted, on the basis of the scores, into one of these four segments. From a table of random numbers, 15 subjects from each range were selected, giving a total of 60 subjects. Thus, the entire range of the distribution was represented.

The 15 subjects in each section were assigned to one of the following three conditions by random procedure: reinforcement of positive self-statements, reinforcement of negative self-statements, and control condition with random reinforcement. Thus, 5 subjects from each Autonomy group were included in each condition, giving a total of 20 subjects in each condition.

The selected subjects were asked to participate in individual interviews. All interviews were conducted by the same interviewer and recorded on tape in the presence of the subject. The subjects were instructed to talk about themselves, covering both positive and negative characteristics. During the first 10 minutes, the interviewer said nothing and made no gestures. This was in order to establish the spontaneous or base rate emission of self-statements for each individual. During the second 10 minutes, the interviewer reinforced the preselected response unit; the reinforcement cue was verbalization of "um-hmm" accompanied by a head nod. The control group received random reinforcement at 10-second intervals during the experimental period, regardless of the response of the subject. The interview was terminated at the end of 20 minutes. The tapes were transcribed with all comments of the interviewer omitted in the transcriptions in order to avoid any influence on the judges.

Judges were two graduate students in clinical psychology. They were asked to rate the interviews for each of the 10-minute intervals as to whether or not a statement was a self-statement and if so, whether it was positive, negative, or neither. They were given explicit criteria for these ratings. The instructions to the judges were taken largely from Raimy (1943). The response unit was defined as a grammatical sentence, and each response unit was numbered.

RESULTS AND DISCUSSION

The obtained reliabilities between judges and for the same judge on two separate occasions were high (91.3% and 97.1%, respectively). In addition, the coder's consistency was found to be high (98%).

Analysis of the data revealed a negative linear correlation of .27, which does not reach significance at a .05 level. Because the direction was as expected, the data were examined more thoroughly to gain a clearer picture of what may be operative in the relation of Autonomy to verbal conditioning effects. The data were reanalyzed according to each type of statement reinforced.

A significant difference ($p < .01$) was found between reinforcement conditions (i.e., whether positive self-statements were reinforced, whether negative self-statements were reinforced, or whether reinforcement was random) in their effects on the proportionate change in positive self-statements. Specifically, the proportion of positive self-statements emitted increased significantly when positive self-statements were reinforced. The relationship of this increase to Autonomy scores ($r = -.60$) is significant beyond the .01 level. A similar increase in the proportion of negative self-statements when these were reinforced was not found. There were, of course, no significant changes in the proportions of either type of statement for the control group under random reinforcement.

Because of the puzzling problem of why there was no significant increase in negative self-statements when they were reinforced, the data were reanalyzed to investigate the possibility that subjects scoring low in Autonomy gave significantly more negative self-statements in their base rates. No significant differences were found in base rate emission of positive or of negative self-statements among Autonomy levels. It is probable that our understanding of Autonomy as a personality variable is inadequate and that some other variable(s) is associated with high or low Autonomy scores, requiring differential predictions of conditioning effects on positive and negative self-statements. This suggestion is consistent with the general expectation that personality variables impose a limitation on conditioning effects.

In a similar study, Rogers (1960) found a significant conditioning effect for reinforced negative self-statements but *not* for reinforced positive self-statements. Clearly, some unspecified variable(s) is exerting a differential effect in these two studies. It may be something as relatively simple as sex differences between experimenters, or it may be something as complex as differences in total setting or "atmosphere" induced. At least, it seems clear that these apparently contradictory results

hardly indicate a straightforward conditioning process!

It is interesting to note that Rogers' control group (no reinforcement) showed a significant *decrease* in positive self-statements but not in negative self-statements. If the effect of no reinforcement is to decrease the subjects' self-statements, then both positive and negative self-statements should decrease. If so, then something else was operative to arrest the decrease in negative self-statements. Since the control group used in the present investigation (receiving random reinforcement) showed no differential effects, the use of random rather than no reinforcement may be a more meaningful control in verbal conditioning studies.

The results obtained from this experiment, then, indicate that by responding selectively to a certain class of verbal behavior, an interviewer can influence the verbalizations of the subject. To be more specific, it was found that the proportion of positive self-statements is increased when they are reinforced and that the degree of the increase is inversely related to scores on Autonomy. It has *not* been demonstrated, however, that verbal conditioning *is* what occurs in therapy, or that conditioning effects are stable over time, or that they are correlated with changes in personality, or that they are related to extra-interview behavior.

Until applied studies of the effects of verbal conditioning are actually carried out, it cannot, of course, be more than conjectured that verbal conditioning techniques bring about the therapeutic changes currently envisioned. Changes in verbal behavior may be necessary concomitants of psychotherapy, but are they either causal or sufficient? The results obtained from the present study in comparison with those obtained by Rogers argue against a straightforward conditioning process.

SUMMARY

The purpose of this experiment was to explore the relationship between a personality variable—Autonomy—and verbal conditioning effects in a quasi-therapy situation when the interviewer employed simple reinforcing behavior. A significant increase in the proportion of emitted positive self-statements was obtained when they were reinforced. The increase in the proportion of positive self-statements when they were reinforced was negatively related to scores on the Autonomy variable, but no similar changes in reinforced negative self-statements were found. Finally, it was suggested that the use of random reinforcement constitutes a more meaningful control in verbal conditioning studies than does the use of no reinforcement.

REFERENCES

EDWARDS, A. L. *Manual: Edwards Personal Preference Schedule.* New York: Psychological Corporation, 1954.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

MOWRER, O. H. Neurosis, psychotherapy, and two-factor learning theory. In O. H. Mowrer (Ed.), *Psychotherapy: Theory and research.* New York: Ronald, 1953.

RAIMY, V. C. Self-reference in counseling interviews. *J. consult. Psychol.*, 1948, **12**, 153–163.

ROGERS, C. R., & DYMOND, ROSALIND F. (Eds.) *Psychotherapy and personality change.* Chicago: Univer. Chicago Press, 1954.

ROGERS, J. M. Operant conditioning in a quasi-therapy setting. *J. abnorm. soc. Psychol.*, 1960, **60**, 247–252.

SALZINGER, K., & PISONI, STEPHANIE. Reinforcement of affect responses of schizophrenics during the clinical interview. *J. abnorm. soc. Psychol.*, 1958, **57**, 84–90.

SARASON, I. G. Interrelationships among individual difference variables, behavior in psychotherapy, and verbal conditioning. *J. abnorm. soc. Psychol.*, 1958, **56**, 339–392.

SHAFFER, L. F. The problem of psychotherapy. *Amer. Psychologist*, 1947, **2**, 459–467.

SHOBEN, E. J. Psychotherapy as a problem in learning theory. *Psychol. Bull.*, 1949, **46**, 366–392.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.

VERPLANCK, W. S. The control of the content of conversation: Reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, **51**, 668–676.

WOLPE, J. *Psychotherapy by reciprocal inhibition.* Stanford, Calif.: Stanford Univer. Press, 1958.

(Received November 30, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 44-54

ATTITUDES TOWARD OLD PEOPLE:
THE DEVELOPMENT OF A SCALE AND AN EXAMINATION OF CORRELATES[1]

NATHAN KOGAN[2]
Educational Testing Service

In recent years there has been considerable discussion regarding the "minority group" status of old people in American society. Barron (1953), for example, has stressed some of the similarities between the position of the aged and of ethnic, racial, and religious minorities on the American scene. In particular, Barron refers to the stereotyping of the aged by other age groups and the discrimination against older people in employment and other areas of life.

More recently, Drake (1958) has pointed to some of the major differences between the status of old people and that of the traditional minorities in the United States. These differences stem from the fact that the aged do not constitute an independently functioning subgroup with a unique history, language, and culture. With these considerations in mind, Drake suggests that old people be designated a "quasi-minority."

Proceeding upon the assumption that old people in American society are devalued, Linden (1957a, 1957b) has outlined various cultural influences considered responsible for such devaluation. Illustrative of these influences, according to Linden, are the diminishing acceptance of family responsibilities towards one's elders, the decline in respect toward the aged as a consequence of loss of their position of authority, and the exaggerated premium placed on the physical and psychological attributes of youth.

Empirical support for a minority or quasi-minority view of old people is offered in a paper by Tuckman and Lorge (1953). They examined the incidence of yes and no responses to a set of statements reflecting stereotypes and misconceptions about the aged in such general areas as physical change, family relationships, personality traits, etc. While this work is of interest as an initial empirical venture in the field, Tuckman and Lorge make no use of attitude scaling procedures and pay little attention to psychological correlates of attitudes toward old people.[3]

The present research is directly concerned with the development of a Likert scale to facilitate the study of attitudes toward old people with respect to both norms and individual differences. The search for correlates of these attitudes proceeded along several directions. First, with a minority group model in mind, we asked whether attitudes toward old people would be related to authoritarianism, anomie, and ethnic prejudice. Second, we considered Barker's (1948) discussion of the minority group aspects of physical disability in our society, and, accordingly, inquired whether attitudes toward old people would be associated with attitudes toward people distinguished by various physical disabilities. Third, since mental deterioration is often alleged to be an important attribute of old age, the relationship between attitudes toward the mentally ill and toward old age was examined. Finally, we hypothesized that there would be significant relationships between attitudes toward old people and such personality dimensions as autonomy, achievement, nurturance, self-esteem, and misanthropy; subjects more favorably disposed toward old people were

[1] This study was supported by the National Institute of Mental Health, United States Public Health Service, under Research Grant M-1867, and was conducted under the auspices of the Age Center of New England, Inc. This work was done in part at the Littauer Statistical Laboratory, Harvard University, and in part at the MIT Computation Center, Cambridge, Massachusetts. A preliminary version of this paper was presented at the April 1959 meetings of the Eastern Psychological Association.

[2] Grateful thanks are due Albert E. Beaton, Arthur S. Couch, and Philip E. Slater for their assistance in machine processing of much of the data. The author also wishes to express his appreciation to Peggy Goldie, Florence C. Shelton, Judith W. Stephens, and Irving K. Zola for their assistance in the preparation of the paper, and to A. Bertrand Warren and Robert Smart for making classes available.

[3] A complete bibliography of this work through 1955 is available in Shock (1957). In a more recent paper, Tuckman and Lorge (1958) concern themselves with personal psychological and physiological symptoms and their projection onto old people as stereotypes.

expected to exhibit stronger tendencies toward self-esteem and nurturance, while subjects more unfavorably disposed were expected to have stronger needs with respect to misanthropy, autonomy, and achievement.

METHOD

Subjects

Students enrolled in introductory classes in psychology served as subjects. Two male samples (N = 128 and 186) were obtained from Northeastern University. One sample (N = 168, 87 males and 81 females) was obtained from Boston University.

Measuring Instruments

A set of 17 items expressing negative sentiments about old people was constructed. A second set of 17 items was then devised the content of which was the reverse of the first set. There were, thus, 17 matched positive-negative pairs. These are shown in Table 1.

Some of the items were adapted from available ethnic minority items by the simple substitution of the "old people" referent. Other statements derive from the author's and others' intuitions regarding stereotypes and feelings about old people in our society. Some a priori clustering of the items is possible on the basis of their manifest content. Thus, Item Pairs 1, 5, and 12 are all concerned with the residential aspects of old people's lives with special reference to segregation, maintenance of home, and character of neighborhood, respectively. Item Pairs 2 and 8 reflect the degree to which vague feelings of discomfort and tension are experienced in the company of old people. The extent to which old people vary among one another is tapped by Item Pairs 11 and 13. The nature of interpersonal relations across age generations—conflicted or benign—is implied in Item Pairs 9, 10, and 16. The theme of dependence is represented by Item Pairs 4 and 17. Item Pairs 3 and 6 refer to the cognitive style and capacity of old people. Qualities of old people with respect to personal appearance and personality are cited in Item Pairs 14 and 15. Finally, Item Pair 7 does not readily cluster with any of the other items.

These items were interspersed among items from other attitude scales thereby partially disguising the presence of logical opposites among the "old people" statements and providing the necessary data for the study of correlates. Most of these additional items were taken from the California F and anti-Negro scales (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950),[4] Srole's (1956) Antiminority and Anomie scales,[5] Gilbert and Levinson's (1956) scale of attitudes toward mental illness[6] and the attitude-to-blindness scale of Cowen, Underberg, and Verrillo (1958).[7] In addition, items measuring attitudes toward totally deaf and crippled persons were constructed, and, in some cases, adapted from the attitude-to-blindness scale cited above.[8] In the case of the items taken from

[4] For the BU and NU-I samples, the following F Scale items (Forms 45 and 40) were used: 1, 6, 9, 18, 21, 25, 26, 27, 34, 37, 42. In addition, the following item was included: "The best teacher or boss is the one who tells us just exactly what is to be done and how to go about it." For the NU-II sample, we employed the five items with the highest loadings on a factorially pure scale of "acceptance of authority" (Bales & Couch, 1956). The scale contains items Nos. 1, 21, and 27 from Forms 45 and 40 of the F Scale, item No. 3 from the Patriotism subscale of the E Scale, and the item "You have to respect authority and when you stop respecting authority, your situation isn't worth much." The following items were selected from the Negro subscale of the E Scale: 5, 8, 11, 14, 16, 22, 28, 31.

[5] Srole's Anomie Scale was expanded on the basis of prior factor analytic work (Zola & Scarr, 1958) to include the following items: "The way the world is today it is hard to find any meaning or purpose in life." "Man is basically evil." "In this world it's dog eat dog." "The forces operating in the world today are so powerful that no single person can do anything to help." In addition, the item "Human nature being what it is, there will always be war and conflict" was added for the NU-II administration. This item, purportedly measuring authoritarianism in the F Scale, proved to be more highly correlated in the present case with anomie items. In general, the correlations in the present samples between the new items and the original Srole statements were as high as the intercorrelations among the latter.

[6] Seven items were selected from this scale on the basis of high DP values and applicability to college student samples. The following items were used: 5, 7, 15, 17, 19, 25, and 31.

[7] The following 10 items were used: 1, 2, 8, 12, 13, 21, 22, 23, 29, and 30.

[8] One-half of the items dealing with totally deaf and crippled persons were constructed by making the appropriate substitution of disability group in the Cowen, Underberg, and Verrillo (1958) items on blindness. The following items from Cowen et al. were converted into cripple items: 3, 10, 14, 17, 19. The following items from Cowen et al. were converted into deafness items: 4, 5, 9, 18, 26. The remaining half of the deafness and cripple items were constructed by the author on the basis of various stereotypes commonly attributed to deaf and crippled persons as discussed by Barker, Wright, Meyerson, and Gonick (1953). The deafness items read as follows: "The peculiar facial expressions of the totally deaf suggest that there is something very queer about them." "I think I would find it intolerable to have to spend a considerable period of time with a totally deaf person." "Most totally deaf persons can be expected to have a rather sour disposition." "The best way to stamp out deafness and other disabilities is to prevent marriages of people with such disabilities." "Totally deaf persons would probably be best off living among themselves away from the world of the hearing."

The cripple items read as follows: "It is very uncomfortable to be with a crippled person because most of them have a chip on their shoulder." "It would be best if crippled people were supported by the government rather than trying to find jobs for them." "Whenever I see a crippled person, I try my best to avoid him." "It would be best if crippled children went to special

TABLE 1

MEANS, STANDARD DEVIATIONS, ITEM-SUM CORRELATIONS (r_{IS}), AND MATCHED ITEM PAIR CORRELATIONS (r_{NP}) FOR "OLD PEOPLE" (OP) ITEMS

Item	BU Sample (N = 168)				NU-I Sample (N = 128)				NU-II Sample (N = 186)			
	M[c]	SD	r_{IS}	r_{NP}	M	SD	r_{IS}	r_{NP}	M	SD	r_{IS}	r_{NP}
1 N[a] It would probably be better if most old people lived in residential units with people of their own age.	3.01	1.54	.40		3.25	1.78	.47		3.34	1.75	.49	
1 P[a] It would probably be better if most old people lived in residential units that also housed younger people.	3.52	1.55	.34	.26	3.53	1.61	.36	.42	3.73	1.67	.32	.32
2 N There is something different about most old people: it's hard to figure out what makes them tick.	2.90	1.40	.57		3.06	1.47	.48		3.01	1.35	.54	
2 P Most old people are really no different from anybody else: they're as easy to understand as younger people.	3.90	1.63	.61	.27	3.72	1.71	.63	.18	3.69	1.56	.57	.33
3 N Most old people get set in their ways and are unable to change.	4.89	1.46	.41		5.18	1.45	.39		5.07	1.49	.49	
3 P Most old people are capable of new adjustments when the situation demands it.	3.96	1.62	.54	.35	3.86	1.77	.33	.17	4.15	1.71	.55	.25
4 N Most old people would prefer to quit work as soon as pensions or their children can support them.	2.08	1.11	.32		2.15	1.17	.10		2.28	1.31	.30	
4 P Most old people would prefer to continue working just as long as they possibly can rather than be dependent on anybody.	2.38	1.13	.28	.38	1.99	.73	.18	.31	2.24	1.07	.32	.35
5 N Most old people tend to let their homes become shabby and unattractive.	2.22	1.15	.47		2.28	1.12	.38		2.43	1.29	.41	
5 P Most old people can generally be counted on to maintain a clean, attractive home.	2.95	1.31	.55	.42	2.69	1.20	.49	.21	2.63	1.18	.59	.32
6 N It is foolish to claim that wisdom comes with old age.	4.05	1.98	.19		3.73	2.00	.10		—	—	—	
6 P People grow wiser with the coming of old age.	3.71	1.49	.26	.40	4.06	1.74	.13	.41	—	—	—	—
7 N Old people have too much power in business and politics.	2.77	1.29	49		3.36	1.67	.41		3.46	1.72	.40	
7 P[b] Old people should have more power in business and politics.	5.15	1.03	.34	.12	5.22	1.19	.36	.13	5.49	1.09	.09	.24
8 N Most old people make one feel ill at ease.	2.82	1.30	.51		3.01	1.32	.50		3.03	1.37	.56	
8 P Most old people are very relaxing to be with.	3.95	1.47	.67	.30	3.65	1.51	.67	.34	3.70	1.49	.62	.44
9 N Most old people bore others by their insistence on talking about the "good old days."	3.64	1.48	.58		3.55	1.56	.53		3.53	1.60	.61	
9 P[b] One of the most interesting and entertaining qualities of most old people is their accounts of their past experiences.	2.79	1.37	.27	.16	2.54	1.30	.23	.17	2.65	1.29	.32	.26
10 N Most old people spend too much time prying into the affairs of others and giving unsought advice.	3.63	1.54	.65		3.41	1.59	.70		3.95	1.57	.67	
10 P[b] Most old people tend to keep to themselves and give advice only when asked.	4.84	1.38	.47	.25	4.74	1.54	.29	.23	4.67	1.51	.41	.26
11 N If old people expect to be liked, their first step is to try to get rid of their irritating faults.	3.82	1.56	.44		3.87	1.69	.48		4.17	1.56	.43	
11 P When you think about it, old people have the same faults as anybody else.	2.60	1.31	.41	.07	2.51	1.25	.49	.14	2.67	1.36	.55	.13
12 N In order to maintain a nice residential neighborhood, it would be best if too many old people did not live in it.	2.37	1.41	.48		2.69	1.65	.25		2.58	1.63	.49	
12 P You can count on finding a nice residential neighborhood when there is a sizeable number of old people living in it.	3.62	1.42	.36	.25	3.46	1.48	.34	.31	3.90	1.51	.34	.06
13 N There are a few exceptions, but in general most old people are pretty much alike.	2.60	1.56	.37		2.74	1.63	.42		3.15	1.80	.44	
13 P It is evident that most old people are very different from one another.	3.63	1.56	.42	.41	3.65	1.74	.38	.41	4.05	1.68	.33	.33
14 N Most old people should be more concerned with their personal appearance; they're too untidy.	3.01	1.36	.44		2.99	1.44	.54		3.08	1.39	.49	

TABLE 1—(*Continued*)

Item	BU Sample (N = 168)				NU-I Sample (N = 128)				NU-II Sample (N = 186)			
	M[c]	SD	r_{IS}	r_{NP}	M	SD	r_{IS}	r_{NP}	M	SD	r_{IS}	r_{NP}
14 P Most old people seem to be quite clean and neat in their personal appearance.	3.54	1.50	.57	.48	3.27	1.48	.56	.44	3.32	1.49	.59	.61
15 N Most old people are irritable, grouchy, and unpleasant.	2.34	.94	.54		2.47	1.22	.46		2.69	1.45	.55	
15 P Most old people are cheerful, agreeable, and good humored.	3.62	1.40	.66	.27	3.35	1.36	.64	.25	3.47	1.51	.61	.40
16 N Most old people are constantly complaining about the behavior of the younger generation.	4.89	1.62	.50		5.18	1.67	.52		5.49	1.46	.54	
16 P One seldom hears old people complaining about the behavior of the younger generation.	5.47	1.13	.49	.41	5.59	1.19	.33	.50	5.75	1.08	.25	.34
17 N Most old people make excessive demands for love and reassurance.	3.96	1.63	.46		3.96	1.68	.50		3.92	1.54	.52	
17 P Most old people need no more love and reassurance than anyone else.	4.54	1.55	.39	.19	4.41	1.79	.27	.27	4.46	1.53	.43	.11
Total negative scale (OP−)	54.87 3.23[d]	11.04			56.84 3.35	11.00			54.17 3.40	12.28		
Total positive scale (OP+)	64.14 3.77	10.90			62.13 3.65	9.70			60.42 3.77	10.47		

Note.—The magnitude of r's required for statistical significance is as follows: for N = 168, .15 at the .05 level and .20 at the .01 level; for N = 128, .17 at the .05 level and .23 at the .01 level; for N = 186, .14 at the .05 level and .19 at the .01 level.

[a] Items are listed in pairs, N representing the negatively worded form and P the positively worded form.

[b] Revised form of the item used with NU-II sample. The revised items read as follows:

7 P Old people have too little power in business and politics.

9 P One of the more interesting qualities of most old people is their accounts of their past experiences.

10 P Most old people respect others' privacy and give advice only when asked.

[c] Negative and positive means were made comparable by subtracting the positive means from 8.00. By this step, higher mean values reflect more unfavorable attitudes for both positive and negative items.

[d] Per item means.

he F Scale and the Gilbert and Levinson mental illness cale, a set of matched items of opposite sign was added o the final battery.[9] Six response categories were provided for all of the items: "strongly disagree," "disagree," "slightly disagree," "slightly agree," "agree," and "strongly agree." These categories were scored 1, 2, 3, 5, 6, and 7, respectively, with a score of 4 assigned in the rare case of failure to respond to an item.

All three samples received the F Scale items, two samples received the Anomie scale, and one sample received all of the remaining scales. Two of the three samples, in addition, were given a set of 30 true-false personality items taken largely from Murray's *Explorations in Personality* (1938) volume, and the Brodbeck and Perlmutter (1954) self-dislike scale.[10]

schools, rather than have them try to keep up with normal children." "I cannot imagine myself having a date with a crippled person."

[9] The F items used with the BU and NU-I samples were selected from Chapman and Campbell (1957)—reversals of Items 6, 9, 18, *21*, *27*, 34, and 44 from Form 40 and Item 44 from Form 60 of the original F Scale—and from Jackson and Messick (1957)—Items *1*, 15, 16, 24, and 28 of their reversed F Scale. For the five-item factorially derived F Scale used with the NU-II sample, the three F items in italics above were used. Reversed wording for the remaining two items was "The only way to have progress in this world is to reject authority when we disagree with it, and to go off on one's own track," and "There are qualities more important than patriotism and loyalty in deciding what makes a good citizen."

The reversed (positively worded) MI items read as follows: "Most people who develop mental disturbances will get over them if treated with kindness, understanding, and sympathy." "The behavior of most mental patients doesn't seem so odd once you understand their childhood and family backgrounds." "A religious person is just as likely to become mentally ill as one who is not religious." "Mental patients deserve the same kind of consideration that is given to patients with physical ailments." "The large majority of discharged mental patients can be expected to make a satisfactory adjustment in their community." One additional MI+ item (No. 7) was selected from Gilbert and Levinson (1956), and in addition, the following MI− item was added: "Once a mental case, always a mental case."

[10] The following 20 questionnaire items were selected from the Murray volume: Achievement items 2, 10; Autonomy items 2, 6; Narcism items 15, 16; Rejection items 1, 3, 4, 10; Nurturance items 3, 5, 6, 7, 9, 11, 13, 14, 16, 20. The 10 items chosen from the Brodbeck and Perlmutter scale were as follows: 2, 5, 9, 10, 11, 14, 16, 20, 23, 26.

RESULTS

Old People Scales

Table 1 presents the basic data for "old people" items and scales (hereafter referred to as OP items and scales). Consider first the scale means shown at the bottom of Table 1. A high degree of consistency among the sample means is evident. In all three samples, for example, the negatively worded statements elicit more favorable sentiments. In other words, subjects disagree more with statements commenting adversely on old people than they agree with statements praising old people. A test for the statistical significance of these overall mean differences yields t's significant beyond the .01 level for each of the samples. Note further that the per item means are uniformly smaller than 4.00—the hypothetical indifference point. These college subjects, then, tend, in general, to be more favorable than unfavorable in their attitudes toward old people as manifested by responses to the OP items of the present battery.

For the BU sample, it was possible to examine sex differences in OP scale means. For both the positive and negative OP scales, no significant differences obtained. This does not imply, of course, that there are no sex differences for specific OP items.

Reliability coefficients and interscale correlations are listed in Table 2. The former (in the diagonals of the table) range from .66 to .85. There is a trend toward greater reliability for the OP− as opposed to the OP+ scale. The magnitude of the OP− reliability coefficient for the NU-II sample suggests that few additional item modifications are required. The OP+ scale, on the other hand, falls somewhat short of an acceptable reliability level, despite item revisions for the NU-II sample.

Consider next the interscale correlations in Table 2. In the construction of matched pairs, we attempted to build genuine logical opposites in terms of connotative meaning rather than retain virtually identical wording in both members of the item pair. This procedure was followed in order to disguise the presence of logical opposites among the items. We wanted to discourage consistency of response based on sheer recognition of almost identically worded opposites.

There is, however, a price paid for such freedom in item wording when subjects endorse or reject two items of presumably opposite meaning. One might account for such findings in terms of response set, ambivalent feelings regarding the object of the attitude, or poorly constructed opposites; or, more likely, a combination of all of these might be involved. This problem has been cogently discussed in the case of F Scale items by Christie, Havel, and Seidenberg (1958).

In the present case, product-moment coefficients between positively and negatively worded scales are uniformly positive. As Table 2 shows, r's between OP+ and OP− range from .46 to .52 in the three samples, all significant beyond the .01 level. Given correlations that can theoretically range from +1.00 (perfect logical consistency) to −1.00 (perfect logical inconsistency), content must necessarily be a more powerful influence than response set for the OP items of the present battery.

TABLE 2

ODD-EVEN RELIABILITY COEFFICIENTS (Spearman-Brown) AND INTERSCALE CORRELATIONS FOR THE OP SCALES

	BU Sample (N = 168)		NU-I Sample (N = 128)		NU-II Sample (N = 186)	
Scale	OP−	OP+	OP−	OP+	OP−	OP+
OP−	.76	.51	.73	.52	.83	.46
OP+		.77		.66		.73

Old People Items

Substantial variations in mean levels across OP items may be noted. It is especially noteworthy that the Item Pair 16 with the highest means—i.e., most unfavorable toward old people—concerns old people's feelings about the younger generation. No single item pair seems to stand out at the favorable end of the attitude continuum. Rather, there are a number of cases where one or both members of an item pair yield means less than 3.00.

Individual items range widely in the extent of their correlation (r_{IS}) with positive and negative sum scores. It will be noted that Item Pair 6 yields very low r_{IS} values for both the BU and NU-I samples and was, accordingly, not used with the NU-II sample. Apparently, the association between wisdom and old age tapped by Item Pair 6 is off the favorability-unfavorability continuum represented

by the other "old people" items. Perhaps, the aphoristic quality of that item pair militates against adequate discriminability.

Item Pair 4 also yields rather low item-sum correlation coefficients. Modification of this item pair is recommended in future work with the OP scales. A suggested modification would eliminate the combination of support by children and by pension in Item 4N. There is an obvious value distinction between support by one's children and support by a pension plan to which one has financially contributed.[11]

An additional consideration in the revision of items is the size of r_{NP}—the correlation between the matched positive and negative items. There is considerable variation among specific OP item pairs in the extent to which the subjects respond to the positive and negative member of the pair in a consistent fashion. Again we may note that such variation could well reflect success in constructing logically opposite item pairs. So, for example, Item Pair 7 does not meet specifications of logical oppositeness, for the negative member of the pair is a descriptive statement while the positive member is a statement advocating change. It is hardly surprising under these circumstances that r_{NP} values in the BU and NU-I samples are nonsignificant. However, when Item 7P was revised for the NU-II sample (cf. Footnote c, Table 1), the r_{NP} value increased, but at the unfortunate price of a sharp drop in the item-sum correlation (r_{IS}). Clearly, further modifications would seem to be in order for Item Pair 7.

Revision of Item 9P was also deemed advisable in view of fairly low r_{NP} (as well as low r_{IS}) values. We suspected that the reference to "entertaining qualities" in the item could be interpreted by subjects in a satirical vein and, accordingly, dropped the presumably offending words in the revised version used with the NU-II sample. A slight increase in r_{IS} and a substantial increase in r_{NP} resulted. It should be noted, however, that r_{IS} for Item 9P is considerably smaller than the corresponding value for its matched negative.

Finally, Item 10P was revised in an effort to make that item more positive in its connotation. In its original form, the item suggests that old people are inclined to avoid interpersonal relations with others. While the revised wording no longer carries this connotation, no appreciable increase in r_{IS} and r_{NP} was observed in the NU-II sample.

In sum, it should be emphasized that the variability in r_{NP} coefficients for specific item pairs is entirely within the positive band of the correlational spectrum. Further, of the 50 r_{NP} values reported in Table 1, 43 are statistically significant at the .05 level or better in the positive, i.e., logically consistent, direction. These results further emphasize that the OP items tap something more than stylistic response tendencies.

Relations with Authoritarianism

We turn next to the various correlates of attitudes toward old people. It is in the examination of correlates that the advantage of using both positively and negatively worded items is especially evident. Several statistically significant relationships obtained between attitudes toward old people and other attitudinal variables based on items worded in the same direction might well be spurious relationships attributable largely to response set tendencies. This phenomenon is best illustrated by the results shown in Table 3.

In each of the samples, statistically significant correlations ranging from .21 to .28 are obtained between F Scale scores for items worded in the authoritarian direction and scores based on negatively worded OP items. It should be noted that these results are

TABLE 3

INTERCORRELATIONS (Product-Moment r's) OF "OLD PEOPLE" SCALES WITH AUTHORITARIANISM AND ANOMIE SCALES

Scale	Sample	"Old People" Scales	
		OP+	OP−
Authoritarianism original (F+)	BU	−.12	.28**
	NU-I	−.16	.21*
	NU-II	−.29**	.21**
Authoritarianism opposites (F−)[a]	BU	.04	.10
	NU-I	.06	.08
	NU-II	−.06	−.22**
Anomie (A)	NU-I	.17*	.39**
	NU-II	.20**	.45**

[a] Scored in the content direction consistent with F+, i.e., higher scores reflect more authoritarianism.

* Significant at the .05 level.

** Significant at the .01 level.

[11] I am grateful to Willis Saulnier of Raytheon Mfg. Co. for this suggestion.

consistent for both a conventional type of F Scale (administered to the BU and NU-I samples) and a "pure" factorially derived scale of "acceptance of authority" (administered to the NU-II sample). The correlations suggest that more authoritarian persons are more *unfavorably* disposed toward old people. When, however, these authoritarianism scores are related to scores based on positively worded "old people" items, the correlation coefficients are in the negative direction, i.e., more authoritarian individuals are more *favorably* disposed toward old people. In the latter case, however, only the r for the NU-II sample is significant.

For items worded in the nonauthoritarian direction, the r's tend toward low nonsignificant values. The significant negative correlation between F− and OP− for the NU-II sample is in the opposite direction and of approximately the same magnitude as the r between F+ and OP− for that sample. There seems little doubt, then, that response set effects partially account for relations between F and OP. Even if F+ items constitute a more valid measure of authoritarianism than do F− items (Gage, Leavitt, & Stone, 1957), the fact remains that F+ correlates in opposite directions with positively and negatively worded OP items. It seems doubtful that the slightly lower reliability of the positive OP scale could account for the extent of this inconsistency.

Correlations between F+ and F− were .42, .27, and .26 for the BU, NU-I, and NU-II samples, respectively. While these r's—all significant beyond the .01 level in the content direction—suggest a reasonable degree of success in the selection of reversed F items, reliabilities for F− (.27, .10, .46) are considerably lower than for F+ (.62, .73, .60) in the BU, NU-I, and NU-II samples, respectively.

Relations with Anomie

Consider next the Anomie (A) scale. As defined operationally by Srole's (1956) items, the anomie variable measures such characteristics as the subjects' pessimism about the future, helplessness in the face of powerful social forces, and inability to find meaning or purpose in life. Here we did not attempt the construction of items of opposite sign. The influence of response set effects may be gauged, however, in terms of the direction and magnitude of the correlations between A and the two OP scales. As Table 3 shows, the r's are considerably higher with the negatively worded OP scale than with the positively worded version in the two samples tested. However, while the r's drop in the latter case, they remain statistically significant in a logically consistent direction. In sum, subjects unfavorably disposed toward old people are more disposed toward anomie.

Is anomie also related to attitudes toward the other minority groups represented in the present study? Correlations between anomie, on the one hand, and scales for attitudes toward ethnic and physically disabled minorities, on the other, range from .28 to .44 in the NU-II sample, all significant beyond the .01 level. Unfortunately, these r's are based on items worded in but one direction, and, hence, are susceptible to the usual effects of response set. Indeed, when anomie is correlated against the two "mental illness" scales, the r of .40 between A and MI− is highly significant, while r between MI+ and A is a mere .04.

Relations with Antiminority Attitudes

Correlations between the OP and the various antiminority scales are listed in Table 4. Since these results are based on a single sample, conclusions drawn from them must necessarily be considered tentative. This is particularly

TABLE 4

INTERCORRELATIONS (Product-Moment r's) OF "OLD PEOPLE" SCALES WITH ANTIMINORITY AND DISABILITY SCALES (NU-II Sample)

Scale	OP+	OP−
Mental illness positive (MI+)[a]	.33	.08
Mental illness negative (MI−)	.11	.46
Anti-Negro (AN)	.21	.46
Antiminority (AM)	.25	.43
Blindness (B)	.28	.52
Deafness (Dx)[b]	.21	.48
Deafness (Dy)[c]	.33	.50
Cripple (Cx)[b]	.23	.53
Cripple (Cy)[c]	.21	.42

Note.—For $N = 186$, r's of .14 and .19 are significant at the .05 and .01 levels, respectively.

[a] Scored in the content direction consistent with MI−, i.e., higher scores reflect more negative attitudes toward the mentally ill.

[b] Adapted from Cowen et al. (1958) blindness items by appropriate substitution of disability group.

[c] Items constructed by the author (cf. Footnote 7 to text).

the case for the various scales in Table 4 original to the present investigation. Here, we simply are dealing with sum scores for items grouped together on the basis of manifest content. The reliabilities of these newly constructed "scales" are surprisingly high, however, given the small number of items they contain. Thus, the Spearman-Brown coefficients for the five-item "deafness" and "cripple" scales range from .51 to .60.

Inspection of Table 4 reveals that the large majority of *r*'s are significant beyond the .01 level. For the "mental illness" scales, highly significant positive correlations are found between items worded in the same direction on the two scales, but *r*'s drop to nonsignificant levels when obtained between scales worded in opposite directions. Table 4 further shows that Srole's Antiminority scale, the California anti-Negro scale, and the scales comprising the various physical disability groupings are all significantly related to the OP scales when response set effects are held constant. While such effects may exert an influence, suggested by the larger *r*'s in the OP negative as opposed to the OP positive column, item content would seem to be of major importance in the present case. In sum, subjects negatively disposed toward old people tend to hold unfavorable attitudes toward ethnic, physical disability, and (possibly) mental illness minorities.

Normative Data for Correlates of the OP Scales[12]

The F+ scale means and *SD*s for the BU and NU-I samples are very similar in magnitude to norms listed in *The Authoritarian Personality* for college student samples. The five-item F Scale used with the NU-II sample seems to elicit stronger authoritarian sentiments. It will be recalled that these items dealt exclusively with "acceptance of authority," and were very possibly less extreme in connotation than many of the F Scale items used with the first two samples.

[12] A one-page table giving the means, standard deviations, and reliability coefficients for correlates of the OP scales has been deposited with the American Documentation Institute. Order Document No. 6691 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

Anomie scores in the present samples are of lower magnitude than authoritarianism scores. Unfortunately, norms for college subjects are not available for the Anomie scale. The various antiminority scales yield consistently low means. In general, subjects are more likely to hold favorable than unfavorable sentiments about the various "minorities" sampled by the items of the present study. It should be noted, in addition, that the means and *SD*s obtained for the AN and B scales correspond quite closely to those reported by Adorno et al. (1950) and Cowen et al. (1958), respectively. All of the "minority" scale per item means are lower than the per item means for the negative OP scale shown in Table 1. Comparison with the positive OP scale is not warranted in view of the uniformly negative direction of the antiminority items.

For the particular items employed in the present study, then, subjects more readily express negative sentiments toward old people than toward members of ethnic, mentally ill, and physically disabled groups. This finding must, of course, be considered highly tentative since items across minority clusters vary in content and wording over and above the difference in minority group referent. Further work along these lines might well take the form of developing a pool of items whose content would be applicable to a wide array of minorities.

Personality Correlates of the OP Scales

It will be recalled that certain explicit predictions were made regarding a number of personality dimensions. A factor analysis of the 30-item true-false personality inventory was carried out, combining the BU and NU-I samples and including sex of the subject as a variable. Seven factors were extracted by the Principal Axes method and these were then rotated to a quartimax solution (Neuhaus & Wrigley, 1954).[13] The first three factors are relatively clear. Factor I has highest loadings on those items involving nurturant concern for others. Factor II yields highest loadings on items expressing feelings of self-doubt in relations with other people. Factor III appears to tap rejective misanthropic sentiments. The remaining factors account for relatively small

[13] A three-page table listing factor loadings (quartimax) for the personality inventory items has been deposited with ADI (see Footnote 12).

portions of the overall variance. Sex of the subject yields the highest loading on Factor IV. Loadings of the other items on this factor are quite low. Factor V involves sentiments of yielding to or resisting others' judgments or wishes. Factor VI shows highest loadings on items expressing doubt about personal achievement. Factor VII accounts for a minute portion of the variance, and, hence, was not considered further.

Factor scores for individuals were derived from the factor loadings and correlated with OP scale scores (Table 5). In general the results are in the predicted direction. Factor I—essentially a nurturance factor—is significantly correlated with OP scale scores. The more favorably subjects are disposed toward old people, the greater their score on Factor I in the nurturant direction. Rejective, misanthropic sentiments tapped by Factor III correlate significantly in the predicted direction for only the positive OP scale, suggesting that the observed relation is partially attributable to response set effects. It should be noted that the factors containing the autonomy and achievement items did not correlate with attitudes toward old people, though this may be a reflection of the small number of items representing those dimensions in the inventory. Finally, there seems to be no simple relation between self-esteem tendencies and feelings about old people.

DISCUSSION AND CONCLUSIONS

That attitudes toward old people are scalable is strongly indicated by the present data. The use of items worded in opposite directions has permitted us to examine the extent to which response set effects permeate the various scales. Correlations between the matched positive and negative "old people" scales are in the content direction, but of moderate size. Compared to the present attitude domain, undergraduate subjects seem to show greater logical consistency in responding to ethnic minority items (Prentice, 1956), but greater inconsistency in regard to authoritarianism (Christie, Havel, & Seidenberg, 1958).

TABLE 5

CORRELATIONS (Product-Moment *r*'s) BETWEEN OP SCALES AND FACTOR SCORES DERIVED FROM PERSONALITY INVENTORY
(N = 296)

Scale	I[a]	II	III[b]	IV	V	VI
OP+	.14*	.06	.18**	−.07	.03	−.09
OP−	.21**	−.04	.07	.03	.10	.05

[a] High scores are in the nonnurturant direction.
[b] High scores are in the misanthropic direction.
* Significant at the .05 level.
** Significant at the .01 level.

Consider next the relationships observed between attitudes toward old people and other attitudinal variables. Of particular interest is the fact that attitudes toward old people are related to feelings of anomie and apparently unrelated to authoritarian tendencies. This result is in marked contrast to the finding that prejudice toward ethnic minorities is associated with both authoritarianism and anomie (Roberts & Rokeach, 1956; Srole, 1956). Hence, while "old people" may constitute a minority, such status does not necessarily render them equivalent to ethnic minorities in their relationship to authoritarianism. This is hardly surprising when one considers the unique features of old people as attitude objects. Old people cannot be categorized in strictly outgroup terms, for most individuals will have old people as members of their families. In addition, old people constitute a minority to which most persons will eventually belong. Given these conditions, one might expect a slight positive association between authoritarianism and favorable attitudes toward old people. Note, however, that old people do not occupy an authoritative position in our society, and, indeed, are often weak, exposed, and highly vulnerable. Given the alleged connection between authoritarianism and contempt for weakness, it would follow that the more authoritarian individuals should be more negatively disposed toward old people. Since no consistent significant relationship was found between the two variables in question, countervailing forces of the kind described may be responsible.

How can we account for the significant relationship between feelings of anomie and attitudes toward old people? Certainly, agreement with anomie items implies that the subject has a precarious and threat-oriented view of life. Old age, in the eyes of such subjects, might well represent that period of life when their pessimistic predictions have been confirmed, i.e., old people have no future,

are helpless, and awaiting death without life's meaning made clear; and, indeed, the views of these subjects have some basis in reality for many old people in American society. It seems reasonable, then, that feelings of anomie and negative attitudes toward old people would be related, for the more anomic subjects would be likely to perceive the aged as symbolic of that period of life when an individual is least able to cope with a hazardous environment.

It will be recalled that the relation between attitudes toward old people and toward mentally ill persons was not entirely free of response set effects. The consistent significant relationships found between attitudes toward old people and toward physically disabled groups, however, suggest that subjects are more likely to associate aging with a state of physical disability as opposed to one of mental illness. It should be noted, however, that the Gilbert and Levinson mental illness scale does not contain any items referring specifically to the mental deterioration associated with senility. Subjects may well consider the latter to have a physical locus, and, hence, to be distinguishable from mental illness conceived more generally. Finally, subjects who are prejudiced against ethnic minorities tend to be negatively disposed toward old people, though, as we have seen, the two sets of attitudes do not seem to have a common base in authoritarian tendencies.

In sum, the present data suggest that there is a general trend for subjects to be positively or negatively disposed toward a wide variety of groups deviating in some respect from a hypothetical norm of similarity to self. Hence, for negatively disposed subjects, members of ethnic minorities, physically disabled persons, old people, and possibly mentally ill people are groups to be avoided or rejected. While parts of this broad attitude complex seem to relate directly to a concept of the authoritarian personality, other parts do not. There is some indication in the present data that Srole's concept of anomie may be most relevant to the formation of a wide range of attitudes, though a definitive statement along these lines must necessarily await the development and application of reversed scales for the many variables in question.

Much of the discussion thus far has been couched in terms of the negative end of the attitude continuum. The observed relation between positive attitudes toward old people and nurturant personality dispositions helps to clarify the opposite end of the attitude dimension under consideration. If persons with strong nurturance needs are attracted to those in need of succorance, old people may be highly appropriate objects in this regard. For there is considerable evidence that dependency is one of the more salient qualities attributed to old people by the younger generation (Golde & Kogan, 1959). While such dependency would arouse positive sentiments in nurturant persons, the dependent state of the older person is likely to arouse ambivalence or conscious hostility in those less inclined in the nurturant direction. Unfortunately, we do not know whether nurturant needs play a part in positive orientations toward other kinds of minorities. It is conceivable, however, that the extent of the relationship would be a function of the prominence of dependency as a characteristic of the minority group in question.

SUMMARY

The present paper reports on the development of a Likert scale for assessing attitudes toward old people and discusses empirical relations found between such attitudes and other attitudinal and personality variables. Three samples of college undergraduates served as subjects.

"Old people" items were constructed in the form of positive-negative pairs, yielding two "old people" (OP) scales: a scale containing items making unfavorable reference to old people and a scale containing matched favorably worded items. Item and scale means and standard deviations are reported, as are item-sum correlations and scale reliabilities. The use of matched positive-negative item pairs permitted a test of the extent to which response set effects permeate the scales. Correlations between positive and negative scales were significant in the direction of logical consistency of response. In addition, all correlations between the positive and negative members of specific item pairs were in the logically consistent direction, the large majority significantly so.

A number of significant relationships observed between attitudes toward old people and other attitudinal and personality variables

proved to be partially attributable to response set effects. We were forced to conclude on this basis that no clear consistent relation obtained between authoritarianism as measured by the F Scale and attitudes toward old people. On the other hand, unfavorable attitudes toward old people were associated with feelings of anomie, and with negative dispositions toward ethnic minorities and a variety of physically disabled groups. A nurturance factor derived from a brief personality inventory given to the subjects was significantly correlated with OP scale scores, the more nurturant subjects being more positively disposed toward old people. Theoretical interpretations of observed associations between the OP scales and other measures are advanced in the body of the paper.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

BALES, R. F., & COUCH, A. A factor analysis of values. Unpublished manuscript, Laboratory of Social Relations, Harvard University, 1956.

BARKER, R. G. The social psychology of physical disability. *J. soc. Issues*, 1948, **4**, 28–38.

BARKER, R. G., WRIGHT, BEATRICE A., MEYERSON, L., & GONICK, MOLLIE R. *Adjustment to physical handicap and illness: A survey of the social psychology of physique and disability*. (Bulletin 55, revised) New York: Social Science Research Council, 1953.

BARRON, M. L. Minority group characteristics of the aged in American society. *J. Geront.*, 1953, **8**, 477–482.

BRODBECK, A. J., & PERLMUTTER, H. V. Self-dislike as a determinant of marked ingroup-outgroup preferences. *J. Psychol.*, 1954, **38**, 271–280.

CHAPMAN, L. J., & CAMPBELL, D. T. Response set in the F Scale. *J. abnorm. soc. Psychol.*, 1957, **54**, 129–132.

CHRISTIE, R., HAVEL, JOAN, & SEIDENBERG, B. Is the F Scale irreversible? *J. abnorm. soc. Psychol.*, 1958, **56**, 143–159.

COWEN, E. L., UNDERBERG, RITA P., & VERRILLO, R. T. The development and testing of an attitude to blindness scale. *J. soc. Psychol.*, 1958, **48**, 297–304.

DRAKE, J. T. *The aged in American society*. New York: Ronald, 1958.

GAGE, N. L., LEAVITT, G. S., & STONE, G. C. The psychological meaning of acquiescence set for authoritarianism. *J. abnorm. soc. Psychol.*, 1957, **55**, 98–103.

GILBERT, DORIS C., & LEVINSON, D. J. Ideology, personality, and institutional policy in the mental hospital. *J. abnorm. soc. Psychol.*, 1956, **53**, 263–271.

GOLDE, PEGGY, & KOGAN, N. A sentence completion procedure for assessing attitudes toward old people. *J. Geront.*, 1959, **14**, 355–363.

JACKSON, D. N., & MESSICK, S. J. A note on "ethnocentrism" and acquiescent response sets. *J. abnorm. soc. Psychol.*, 1957, **54**, 132–134.

LINDEN, M. E. Effects of social attitudes on the mental health of the aging. *Geriatrics*, 1957, **12**, 109–114. (a)

LINDEN, M. E. Relationship between social attitudes toward aging and the delinquencies of youth. *Amer. J. Psychiat.*, 1957, **114**, 444–448. (b)

MURRAY, H. A. *Explorations in personality*. New York: Oxford, 1938.

NEUHAUS, J. O., & WRIGLEY, C. The quartimax method: An analytical approach to orthogonal simple structure. *Brit. J. statist. Psychol.*, 1954, **7**, 81–91.

PRENTICE, N. M. The comparability of positive and negative items in scales of ethnic prejudice. *J. abnorm. soc. Psychol.*, 1956, **52**, 420–421.

ROBERTS, A. H., & ROKEACH, M. Anomie, authoritarianism, and prejudice: A replication. *Amer. J. Sociol.*, 1956, **61**, 355–358.

SHOCK, N. W. *A classified bibliography of gerontology and geriatrics*. (Supplement I, 1949–1955) Stanford, Calif.: Stanford Univer. Press, 1957.

SROLE, L. Social integration and certain corollaries: An exploratory study. *Amer. sociol. Rev.*, 1956, **21**, 709–716.

TUCKMAN, J., & LORGE, I. Attitudes toward old people. *J. soc. Psychol.*, 1953, **37**, 249–260.

TUCKMAN, J., & LORGE, I. The projection of personal symptom into stereotype about aging. *J. Geront.*, 1958, **13**, 70–73.

ZOLA, I. K., & SCARR, H. A. Attitude dimensions in older people. Progress Report, October 1958 NIMH Project M-1402. (Available from the Age Center of New England, Inc.; 160 Commonwealth Ave.; Boston 16, Mass.)

(Received November 30, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 55–61

THE RELATIONSHIP BETWEEN MEDICAL AND PSYCHIATRIC SYMPTOMS

RUTH G. MATARAZZO, JOSEPH D. MATARAZZO, AND GEORGE SASLOW
University of Oregon Medical School

It has been observed that number and kind of medical symptoms, as reported by the patient, are related to patient characteristics other than simply his diagnosed medical, surgical, neurological, etc. disease or diseases. Thus, in developing the Cornell Medical Index (CMI) (Brodman, Erdmann, Lorge, & Wolff, 1949, 1953), its authors found that older people check more physical symptoms than do younger people, and that women check more than men. An observation similar to the latter has been made independently (Boake, 1958) in a study of "common respiratory diseases." Boake found that the females in his groups actually had slightly more illnesses on the average than the males, and also that the females had, on the average, slightly more symptoms per illness than the males. He did not, however, present the statistical significance of these differences.

The authors of the CMI also suggest that when a large number of positive replies is made on their Health Questionnaire (including both medical and psychiatric symptoms), psychiatric disturbance is indicated, as is the case when certain replies fall specifically in the "fatigability" and "frequency of illness" categories. Thus, they found that emotional state or adjustment pattern are other patient characteristics which are associated with medical symptomatology (Brodman, Erdmann, Lorge, Gershenson, & Wolff, 1952).

Hinkle and Wolff (1957) have studied further the hypothesis that medical and psychiatric symptoms tend to occur together. They state: "... 'Illness' is a state of the organism and ... when a human moves from a state of 'health' into a state of 'illness,' the 'illness' is likely to be manifested by a variety of syndromes appearing concurrently or consecutively, their nature being dependent upon the various factors acting upon the organism at that time." In studying long-term health records they found that:

1. Individuals having the greatest number of illnesses experienced a wide variety of them (both major and minor) in a number of body systems. There was a relationship between number of major illnesses and number of minor illnesses experienced by an individual, and between number of illnesses and number of body systems involved.

2. Both in those individuals having many illnesses and those having few, illnesses were not scattered randomly through their lives, but usually appeared in clusters.

3. In general, all illnesses were likely to occur in periods when the individual was having difficulty in adapting to the total configuration of his life.

4. A large percentage of those with high susceptibility to illness had had a difficult early life, although this relationship was by no means absolute.

While clusters of illness often could be related to adverse emotional circumstances, there were also periods of emotional stress which apparently were not associated with any significant somatic disturbances. No statistical presentation of their data was made, however.

On the basis of all the above findings (Boake, 1958; Brodman et al., 1952, 1953; Hinkle & Wolff, 1957), then, one might expect to find that:

1. Psychiatric patients, presumably under more intense, longer-lasting, and more frequent stress, in general, than medical patients, would have a greater number of bodily symptoms than would medical patients.

2. The number of psychiatric symptoms (degree of emotional distress) on various tests would be highly correlated with number of "physical" symptoms in both medical and psychiatric groups, or that total number of physical symptoms is nearly as good a measure of psychiatric status as is number of psychiatric symptoms (although no assumption is made as to which may be precursor of the other).

3. The total number of symptoms should be positively correlated with age.

4. Within both groups, women should react to the physical or emotional stress with a

greater total number of symptoms than the men.

Each of these hypotheses is either an independent check on the actual findings or an implication derived from the results of the investigators cited.

METHOD

The 162 subjects used here consisted of 40 medical inpatients, 42 medical outpatients, 40 psychiatric inpatients, and 40 psychiatric outpatients (all white) who were under treatment at the University of Oregon Medical School Hospitals and Clinics. They were taken at random from among those coming through the services at the time the study was being carried out, provided they were willing to fill out the forms, had the education and intelligence to do so, and, for the medical inpatient group, were not too physically incapacitated.

The patients were requested to fill out the test forms used in this research "in order to provide background information on their cases for the doctors." Patients were asked to complete the CMI (long form), the Taylor Manifest Anxiety scale, and the Saslow Psychosomatic Inventory. The CMI is a 195-item questionnaire designed to obtain a standardized medical history (Brodman et al., 1949). It covers, separately, the different body systems; e.g., eye and ear, respiratory, cardiovascular, digestive, musculoskeletal, skin, nervous system, and genitourinary; as well as frequency of illness, miscellaneous diseases, and mood and feeling patterns. Sections A through K cover primarily physical manifestations in separate body systems, whereas L through R cover habits, and mood and feeling patterns. Thus, the total Cornell score was obtained and then broken down into a score for A–K (medical symptoms) and L–R (emotional symptoms).

The *MA* (Taylor, 1953) is a 50-item questionnaire covering both psychological and physical symptoms of anxiety or "neuroticism." It has been found to differentiate psychiatric from medical and nonpatient groups (Matarazzo, Guze, & Matarazzo, 1955).

The Saslow Screening Inventory (Gleser & Ulett, 1952; Saslow, Counts, & DuBois, 1951) is a measure of psychological and physical reactions to anger and to anxiety producing situations.

RESULTS

Analyses of variance were computed in order to ascertain comparability of the four samples in regard to age, which has been reported to be associated with symptomatology (Brodman et al., 1953). It was found that the medical inpatients were significantly older than the medical outpatients (48.35 vs. 27.76, p of .001),[1] and this was true for both males and females. This difference of 20 years on the average is considerable and probably reflects the general fact that the illnesses requiring hospitalization are usually encountered in older age groups. There is no such difference between the psychiatric in and outpatients, but the total psychiatric group is significantly younger than the medical group (33.05 vs. 37.80, p of .05). This difference is due, however, to the advanced age of the medical inpatients, while the medical outpatients do not differ in age from the psychiatric groups. If increased age is associated with more symptoms, this age discrepancy should work against the hypothesis that the psychiatric population will have more medical symptoms than the medical group.

It was also important to determine whether there was an age difference between the sexes because this might complicate the relationship between sex and number of symptoms. Within each of the four groups there were no statistically significant sex differences in age, although there was a consistent trend for the male medical patients to be older than the female medical patients. When the two medical groups (inpatient and outpatient) were combined, this difference was significant (43.71 vs. 33.40, p of .05). There was no sex difference in age within the psychiatric groups. Thus, in this population, it appears that medical treatment is likely to be required or sought at an earlier age in females than in males, generally speaking. There was no such sex difference in psychiatric illness. Since the age difference between the sexes in the medical groups was found to be approximately the same (8 years) for both in- and outpatients, it appears that both minor and major ailments were more prevalent among females who were still in their prime. This age difference, however, should also work against the hypothesis that the males will have fewer symptoms than females, since the former were older.

There were no significant educational differences between sexes, nor was level of education associated with in- or outpatient status. The psychiatric group was, however, slightly superior in education to the medical group (11.35 years vs. 10.00 years, p of .01). This was largely due to the superiority of the psychiatric inpatient sample, which had a mean of 12.00 years of education.

[1] The age ranges for the four groups were as follows: medical outpatients 18–52, medical inpatients 14–68, psychiatric outpatients 16–65, psychiatric inpatients 15–78.

Table 1 presents the means of all groups on each of the scales, and the *F* values for differences among them, and indicates that four out of five measures significantly differentiate the groups, the psychiatric patients having the greater number of symptoms. The one measure which was not significant was the Cornell A–K (medical) items, while the Total Cornell and all the psychiatric measures attained statistical significance. Thus, number of medical symptoms does not discriminate among the groups as a whole. If one were to rank the groups on each of the scores, however, the medical outpatient group would, in every case but one (Cornell L–R), appear to be the lowest on all medical and psychiatric scales. Similarly, the medical inpatient group is, in almost every case, the next in rank, followed by the psychiatric inpatients and psychiatric outpatients, in that order. Thus, one may conclude that, in general, the medical outpatients presumably experienced, and for this reason reported, the least number of medical and psychiatric symptoms (and, thus, inferentially the least amount of stress), while the psychiatric outpatients suffered the most. One might ordinarily have expected the psychiatric inpatient sample to report the greatest number of symptoms, and there is no obvious explanation for the fact that they did not. However, the differences in means between this group and its outpatient counterpart is small on all measures. In regard to medical symptoms (A–K), there is a nonsignificant trend for the four groups to be differentiated, which, however, follows the identical pattern of all the psychiatric scales. All of the other scales differentiate the groups, three of them at the .001 level (Cornell L–R, Anxiety scale, and Psychosomatic Inventory). Analysis by *t* tests established that, on these scales, each of the medical groups was significantly different from each of the psychiatric groups, but differences *within* medical and psychiatric groups were not significant. On the "medical" part of the Cornell Index (A–K), only the two extreme groups were statistically differentiated (p of .05)—the medical outpatients from the psychiatric outpatients. This discrimination, and the consistent trend for medical symptoms to differentiate the groups in the same way as did emotional symptoms, gave some support to Hypothesis 1: the psychiatric patients, presumably under greater total stress than the medical patients, experienced many more psychiatric symptoms and had a tendency toward more physical symptoms. It would appear from Table 1, however, that the number of psychiatric symptoms was a better means of establishing psychiatric status than was the number of physical symptoms.

Table 2 presents the Pearson product-moment correlations of all the measures for the total medical group, including the correlations with age. (Since the comparable correlations for the psychiatric sample indicated that the relationships were identical for both groups, the additional table has been omitted.) From Table 2 it can be seen, first of all, that the total CMI was highly correlated with both of its subparts (.94 with the "medical" A–K and .82 with the "psychiatric" L–R). While it was not surprising to find that the whole was highly correlated with its subparts, the total CMI was almost as highly correlated with the *MA* (.78) and very significantly correlated with the Saslow Psychosomatic Inventory (.55). Furthermore, when only the A–K (medical) scale was used, its correlations with these same psychiatric measures dropped only slightly below those of the total index (.59 with L–R, .62 with the

TABLE 1

COMPARISON OF MEAN SCORES AMONG THE FOUR GROUPS (Sexes Combined)
(N = 162)

	Total Cornell Index	A-K (Medical) Cornell	L-R (Psychiatric) Cornell	Taylor Manifest Anxiety scale	Saslow Psychosomatic Inventory
Medical Inpatients	38.98	29.30	9.68	17.05	6.68
Medical Outpatients	33.50	23.50	10.00	14.81	5.50
Psychiatric Inpatients	48.88	29.86	19.02	25.68	13.18
Psychiatric Outpatients	52.82	31.37	21.45	26.31	12.38
F	4.81	1.50	13.72	11.77	10.44
p level	.05		.001	.001	.001

TABLE 2

INTERCORRELATIONS AMONG AGE AND SCORES FOR TOTAL MEDICAL GROUP (N = 82)

	Total Cornell Index	A-K (Medical) Cornell	L-R (Psychiatric) Cornell	Taylor Manifest Anxiety scale	Saslow Psychosomatic Inventory
Age	.14	.17	.02	.17	.05
Total Cornell Index		.94	.82	.78	.55
A-K (Medical) Cornell			.59	.62	.40
L-R (Psychiatric) Cornell				.82	.61
Taylor Manifest Anxiety scale					.63
Saslow Psychosomatic Inventory					
p .05–.22					
p .01–.28					

Anxiety scale, and .40 with the Psychosomatic Inventory). Thus, the number of medical symptoms was highly related to each of the psychiatric scores. Hypothesis 2, then, was well substantiated; one can, with fair success, predict number of psychiatric symptoms from knowledge of the number of medical symptoms reported and vice versa. This is true, to a greater or lesser extent, depending upon the psychiatric scale used. The lower correlation of the psychosomatic inventory was probably associated with its smaller range of scores.

The remaining correlations, ranging from .61 to .82, show the extent to which the psychiatric measures were correlated, indicating that there is high predictability from one to the other although there is little similarity among the items of which they are comprised. Interestingly, as shown in Table 2, age was found not be associated with number of symptoms of any kind.

Table 3 presents a comparison of male and female scores on all measures within each of the four subgroups. Within both medical subgroups, comparison of the males and females indicated that there were no differences in number of either medical or psychiatric symptoms. Thus, contrary to the results of Brodman et al. (1953) and Boake (1958), the medically sick but presumably emotionally healthy women of this study did not indicate that they had a significantly greater number of medical ailments or concerns than did the men. It should be mentioned, nevertheless, that while no difference between groups attained significance, there was a consistent trend for the women to score higher on each scale, in agreement with the results of these investigators.

TABLE 3

COMPARISON OF MEAN SCORES OF MALES AND FEMALES WITHIN EACH OF THE FOUR GROUPS

	Total Cornell Index	A-K (Medical) Cornell	L-R (Psychiatric) Cornell	Taylor Manifest Anxiety scale	Saslow Psychosomatic Inventory
Medical Inpatients (N = 40)					
Males	35.70	28.30	7.40	15.00	5.05
Females	42.25	30.30	11.95	19.10	8.30
F	.82	.03	3.03	1.81	2.84
Medical Outpatients (N = 42)					
Males	31.07	21.07	10.00	14.53	4.87
Females	34.85	24.85	10.00	14.96	5.85
F	.28	.61	.00	.01	.48
Psychiatric Inpatients (N = 40)					
Males	38.61	23.00	15.61	24.08	8.94
Females	57.27	35.45	21.82	26.88	16.64
F	4.85*	4.03	3.25	.36	7.65**
Psychiatric Outpatients (N = 40)					
Males	41.78	22.78	19.00	26.65	9.61
Females	61.86	38.41	23.45	26.04	14.64
F	4.91*	7.05*	1.32	.02	2.90

* .05 level of probability.
** .01 level of probability.

Among the psychiatric groups, however, the women scored significantly higher on certain scales. Thus, in both psychiatric groups, the women obtained a higher total CMI score than did the men (p of .05). When the CMI was broken down into its "medical" and "psychiatric" subparts, it is apparent from Table 3 that, while there was a trend for the women to score higher than the men

in both categories, it was mainly the "medical" category which differentiated them. This difference attained significance (p of .05) in the outpatient sample, and nearly reached this level for the inpatient group. While the two remaining psychiatric scales showed a consistent trend in both groups for the women to be higher, this difference reached significance only in the Saslow Psychosomatic Inventory for the inpatient group (p of .001). Thus, among the psychiatric patients, the women did have significantly more medical complaints than the men, as indicated by the Cornell A–K score, but also more complaints on the Saslow Psychosomatic Inventory. There was also, as in the medical group, a consistent trend for the women to report more psychiatric symptoms.

In Table 4, a comparison of the four subgroups is presented with the sexes separated. When only the males were compared, the difference (presented in Table 1) among groups on the Total Cornell Index no longer reached significance. In addition, the psychiatric scales, while still discriminating the groups, tended to be less efficient in their power to differentiate the samples.

On the other hand (as seen in Table 4), when the females were compared, the difference between groups on the Total Cornell Index was increased from p of .05 (Table 1) to p of .01, and this greater difference was largely contributed by the medical (A–K) items. While A–K did not differentiate the groups when the sexes were combined (Table 1), it did differentiate the females of the groups (p of .05). For the females, the other (psychiatric) measures were not essentially different in their level of statistical significance from those presented in Table 1 for the two sexes combined. It can be seen (Table 4), however, that in general the females contributed more to the group differences (Table 1) than the males, and that female psychiatric patients differed more from their psychiatrically healthy (medically ill) counterparts than did the males.

DISCUSSION

It is interesting that the females tended to appear for treatment of both minor and major illnesses at an earlier age than did the men (although not for psychiatric treatment), and that the women psychiatric patients differed from all groups of males and from medically ill females in that they reported many more physical (medical) disturbances. An explanation for this discrepancy between the sexes is difficult to find. A greater number of minor complaints could be attributed purely to greater awareness of, or concern with, physical symptoms. However, the same age difference prevailed in both medical outpatient and inpatient groups, the latter presumably having had to convince not only themselves but several physicians that they were seriously ill. (The sex ratio of approximately 1:1 in this

TABLE 4
COMPARISON OF MEAN SCORES AMONG THE FOUR GROUPS

	Total Cornell Index	A-K (Medical) Cornell	L-R (Psychiatric) Cornell	Taylor Manifest Anxiety scale	Saslow Psychosomatic Inventory
Males					
Medical Inpatients	35.70	28.30	7.40	15.00	5.05
Medical Outpatients	31.07	21.07	10.00	14.53	4.87
Psychiatric Inpatients	38.61	23.00	15.61	24.08	8.94
Psychiatric Outpatients	41.78	22.78	19.00	26.65	9.61
F	.61	.66	5.87	6.35	2.98
p level			.01	.001	.05
Females					
Medical Inpatients	42.25	30.30	11.95	19.10	8.30
Medical Outpatients	34.85	24.85	10.00	14.96	5.85
Psychiatric Inpatients	57.27	35.45	21.82	26.88	16.64
Psychiatric Outpatients	61.86	38.41	23.45	26.04	14.64
F	5.51	2.76	8.84	5.95	9.04
p level	.01	.05	.001	.01	.001

study is artificial since an attempt was made to obtain equal numbers of men and women in each group.) It seems probable, then, that women tend to have a somewhat greater number of both major and minor illnesses. Adding weight to this conclusion is the fact that in both psychiatric groups the age difference did not hold. If it were purely a matter of women seeking treatment earlier than men, this should be reflected in the psychiatric groups also. Again, consistent with the hypothesis of greater physical vulnerability among females, was the tendency for the latter to report more symptoms under emotional stress (psychiatric group). Thus, under the same emotional stress (as measured by scores on psychiatric scales), the women reported more physical distress than the men. Again, if it were purely a matter of greater awareness of discomfort, the women would be expected to report significantly more psychiatric symptoms also.

A paper by Mandler, Mandler, and Uviller (1958) is interesting in this regard. They studied a group of subjects which reported high awareness of autonomic changes under stress, and one which reported low awareness. Actual physiological measurements taken in a stress situation tended to corroborate the subjects' reports; i.e., those who reported greater awareness of autonomic changes were those who, by measurement, were found to have the larger physiological responses. Also notable, however, was the fact that the subjects who were most physiologically responsive tended to overestimate their physiological changes, whereas those with low responsiveness tended to underestimate physiological changes. It is possible, then, that women are slightly more sensitive or more physiologically reactive to stress than men, but that they also tend to overestimate the extent of the disturbance, or to be more highly aware of it. This could be analogous to the apparent "conditioning" process involved in many psychological disturbances, such as the anxiety neurotic's "fear of fear." That is, after experiencing episodes of a particular kind of discomfort, one can learn to anticipate or dread further instances of discomfort while becoming progressively sensitive to lower magnitudes of pertinent cues (the "reaction-sensitization" of Cameron, 1947). That, for example, the anxiety neurotic's fear of the symptoms of fear may be at least partly determined by physiological preconditions seems indicated by the report of Cohen (1949) that the vital capacity of the anxious person tends to be lower than that of normals, and that, thus, his breathing, heart rate, etc. may be more vulnerable to stress. Lacey (1956), in demonstrating that each individual has a specific autonomic pattern of reactivity to stress, has provided the possibility of explaining the anxious patient's fear of, or concentration upon, specific physiological symptoms; e.g., one patient's fear that he has heart trouble; another's fear of vomiting, being unable to get his breath, or fainting; another's obsession with headache, constant fatigue, or bowel function. The results of Grossman and Greenberg (1957) suggest that the autonomic patterns described by Lacey are present in the newborn, and, thus, are apparently constitutional, at least in part. That physiological precondition may be an important factor in "psychiatric" symptomatology is further suggested by a report on bronchial asthma, allergic rhinitis, and pulmonary emphysema (Heiskell, Rhodes, & Thayer, 1959). The patients were rated as to severity of functional (breathing) impairment and compared on the Hypochondriasis, Depression, and Hysteria scales of the MMPI. Whereas asthma has been said to be a "psychosomatic" disease, it was found that disease entity did not seem to account for differences in MMPI score, although severity of functional impairment (in all three diseases) did. Thus, the physiological changes, symptoms, or impairments associated with the illness seemed to be related to emotional distress as reflected in the MMPI.

Regardless of etiology, however, the high positive correlation found in the present study between number of medical and psychiatric symptoms (Table 2) tends to confirm the Hinkle and Wolff observation that the state of health or illness is a condition of the organism as a whole, and that when the organism is under stress, it is likely to reflect this stress by symptoms of malfunction in a number of areas.

Summary

This investigation has attempted to relate number of symptoms to patient characteristics other than diagnosed medical, neurological, etc. disease or diseases. Four groups of patients

were studied: psychiatric inpatient, psychiatric outpatient, medical inpatient, and medical outpatient samples, comprised of approximately equal numbers of males and females. It was found that, when the groups were compared for number of symptoms on the Total Cornell Index, Cornell A–K (medical symptoms), Cornell L–R (psychiatric symptoms), Taylor Manifest Anxiety scale, and Saslow Psychosomatic Screening Inventory, in general, the medical outpatients had the fewest symptoms, followed by the medical inpatients, psychiatric inpatients, and psychiatric outpatients, in that order. When medical and psychiatric symptoms were correlated, it was found that number of medical symptoms (A–K) correlated .59 with (L–R) psychiatric symptoms, .62 with score on the anxiety scale, and .40 with number of symptoms on the psychosomatic inventory. Thus, number of medical symptoms was a good predictor of psychiatric symptomatology and vice versa. Female psychiatric patients were found to have many more medical symptoms than either male psychiatric patients or male or female medical patients. Males and females did not differ significantly in regard to number of psychiatric symptoms, however. This was considered in relation to the fact that the female medical patients were approximately 10 years younger than the male medical patients, although there was no such difference in the psychiatric group. The results suggested the hypothesis that women are physically more reactive to stress, and that this physical sensitivity may then "condition" them to greater awareness of, or concern with, the physical manifestations of stress.

In addition, the high relationship between medical and psychiatric symptoms was discussed in relation to individual autonomic response patterns described by Lacey, and to Hinkle and Wolff's hypothesis that the state of health or illness is a condition of the organism as a whole.

REFERENCES

BOAKE, W. C. A study of illness in a group of Cleveland families: XVIII. Tobacco smoking an ' respiratory infections. *New England J. Med.*, 1958, **259**, 1245–1249.

BRODMAN, K., ERDMANN, A., JR., LORGE, I., GERSHENSON, C. P., & WOLFF, H. G. The Cornell Medical Index–Health Questionnaire: III. The evaluation of emotional disturbances. *J. clin. Psychol.*, 1952, **8**, 119–124.

BRODMAN, K., ERDMANN, A., JR., LORGE, I., & WOLFF, H. G. The Cornell Medical Index, an adjunct to medical interview. *J. Amer. Med. Ass.*, 1949, **140**, 530–534.

BRODMAN, K., ERDMANN, A., JR., LORGE, I., & WOLFF, H. G. The Cornell Medical Index–Health Questionnaire: VI. The relation of patients' complaints to age, sex, race, and education. *J. Geront.*, 1953, **8**, 339–342.

CAMERON, N. *The psychology of behavior disorders*. Boston: Houghton Mifflin, 1947.

COHEN, M. E. Neurocirculatory asthenia. *Med. Clin. N. Amer.*, 1949, Summer, 1343–1364.

GLESER, GOLDINE, & ULETT, G. A. The Saslow Screening Test as a measure of anxiety-proneness. *J. clin. Psychol.*, 1952, **8**, 279–283.

GROSSMAN, H. J., & GREENBERG, N. H. Psychosomatic differentiations in infancy: I. Autonomic activity in the newborn. *Psychosom. Med.*, 1957, **19**, 293–306.

HEISKELL, C. L., RHODES, J. M., & THAYER, K. H. Some psychosomatic aspects of asthma. *J. Amer. Med. Ass.*, 1959, **170**, 1764–1767.

HINKLE, L. E., & WOLFF, H. G. Health and the social environment: Experimental investigations. In A. H. Leighton, J. A. Clausen, & R. N. Wilson (Eds.), *Explorations in social psychiatry*. New York: Basic Books, 1957. Pp. 105–137.

LACEY, J. I. The evaluation of autonomic responses: Toward a general solution. *Ann. NY Acad. Sci.*, 1956, **67**, 123–164.

MANDLER, G., MANDLER, J. M., & UVILLER, E. T. Autonomic feedback: The perception of autonomic activity. *J. abnorm. soc. Psychol.*, 1958, **56**, 367–373.

MATARAZZO, J. D., GUZE, S. B., & MATARAZZO, RUTH G. An approach to the validity of the Taylor anxiety scale: Scores of medical and psychiatric patients. *J. abnorm. soc. Psychol.* 1955, **51**, 276–280.

SASLOW, G., COUNTS, R. N., & DUBOIS, P. H. Evaluation of a new psychiatric screening test. *Psychosom. Med.*, 1951, **13**, 242–253.

TAYLOR, JANET, A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

(Received December 2, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 62-73

ANALYSIS OF THE CONNOTATIVE MEANINGS OF A VARIETY OF HUMAN VALUES AS EXPRESSED BY AMERICAN COLLEGE STUDENTS[1]

CHARLES E. OSGOOD, EDWARD E. WARE
University of Illinois
AND CHARLES MORRIS
University of Florida

HUMAN beings display a variety of conceptions of "the good life." Some lay stress on the overcoming of obstacles through persistent striving; some favor the search for sensual comforts and pleasures; others emphasize the more intellectual joys to be obtained in passive contemplation and reflection. Whereas some people perceive values in terms of society, through participation with others toward group goals, others organize their value systems around the essential dignity and independence of the individual. Whether they be expressed explicitly in the doctrines of religion and the policies of political organization, or expressed implicitly in the everyday behaviors of ordinary people living their normal lives, such values are both deeply ingrained and pervasively influential in determining choices, both great and small. As Kluckhohn et al. (1951) put it, "a value-orientation (is) a generalized and organized conception, influencing behavior, of nature, of man's place in it, of man's relation to man, and of the desirable and nondesirable as they may relate to man-environment and interhuman relations" (p. 411). Although the preferential ordering of values probably varies most notably across different groups of men, e.g., across different cultures, it also varies among the individuals composing such groups—for example, differences in the assessment of human values are implicit in the policies favored by different Americans for survival in this nuclear age.

In *Paths of Life* (1956) Morris distinguished three basic components of personality on philosophical grounds, these being presumed to combine in various ways to produce various value orientations: "Dionysian" (tendency to release and indulge existing desires), "Promethean" (tendency to manipulate and remake the world), and "Buddhistic" (tendency to regulate the self by holding in check its desires). The combination of these tendencies in their possible orderings, including that combination in which all are moderately strong (Way 7), generated the 7 "ways to live" which were described and explored in *Paths of Life*. These are the first 7 Ways given in essence in Table 1. When single-paragraph descriptions of these "ways to live" were given to students who were asked to indicate their preferences about them, the need for several additional alternatives was demonstrated. Ways 8 through 13 were added, and this 13-alternative form is the one which was used by Jones and Morris (1956; Morris & Jones, 1955) in a series of empirical studies of human values. These and other studies, with their implications, are summarized in *Varieties of Human Value* (Morris, 1956b). Table 1 attempts to give only the essential emphasis of these Ways, for purposes of identification throughout this report; the more complete descriptive paragraphs used as bases for judgments by subjects in Morris' original studies and in the present investigation may be found in *Varieties of Human Value* (Morris, 1956b, pp. 15–19) and in a paper by Morris and Jones (1955, pp. 524–525).

Morris obtained preferential judgments ("Indicate . . . how much you yourself like or dislike each of them") of the 13 ways, using a seven-step scale from "like very much" to "dislike very much." Sizable samples of American, Indian (Asian), and Chinese college students, and smaller samples in other countries, were used. When the preferential judgments were subjected to factor analysis (Morris & Jones, 1955), evidence for considerable agreement in factor structure across

[1] The basic data used in this study were collected by Olga McNemar in 1957. Norma Mae Ross helped work on this material. I. Richard Savage, Lyle V. Jones, and George J. Suci made helpful suggestions at various stages of the work.

TABLE 1
A SAMPLE OF HUMAN VALUES

Way 1	preserve the best in society
Way 2	self-sufficiency, reflection, and meditation
Way 3	altruistic affection and concern for others
Way 4	abandonment to sensuous enjoyment
Way 5	group action toward common goals
Way 6	progress through realistic solution of problems
Way 7	integration of diversity
Way 8	wholesome enjoyment of simple comforts
Way 9	quiet receptivity to experience
Way 10	dignified self-control
Way 11	contemplation of the rich inner life
Way 12	dynamic physical interaction with the environment
Way 13	humble obedience to cosmic purposes

Note.—These phrases attempt to epitomize the statements of the Ways given in *Varieties of Human Value*, pp. 15-19; they are given to provide easy identification of the Ways, but are not the statements the subjects judged.

different cultures was obtained, even though the preferences for various "ways to live" differed in these cultures. *Varieties of Human Value* describes relationships between these value factors and a number of other types of individual differences, including differences in religious affiliation, in social caste and class, in temperament, in personality, and in body size and build. The fact that such relationships do exist both testifies to the validity of the value dimensions and serves to illuminate some of the ways in which values relate to other aspects of human existence.

For a number of years Osgood and his associates have been working on a psychological theory of meaning and the objective measurement of certain (connotative) aspects of meaning. Application of factor analytic techniques to large numbers of descriptive dimensions, defined by polar adjectives, has provided evidence for the existence of a limited number of connotative semantic factors—"evaluation," "potency," and "activity" being most clearly defined, and "stability" and "receptivity" being less well defined. The logic of this type of measurement, the factorial studies, and applications of the technique to a variety of fields have been described in *The Measurement of Meaning* (Osgood, Suci, & Tannenbaum, 1957). A major purpose of this research has been to make it possible to index the connotative meanings of concepts as locations in a common-dimensional space, and hence to be able to specify similarities and differences in concept meanings as distances (cf. Osgood & Suci, 1952) in this space. Application of this type of measurement has been made to the prediction and measurement of attitude change (Osgood & Tannenbaum, 1955), to a blind analysis of a case of triple personality (Osgood & Luria, 1954), to cross-cultural studies of semantic similarities (Kumata, 1957; Triandis & Osgood, 1958), and to various psycholinguistic, mass communication, and esthetic problems.

The present investigation was planned with the prime purpose of bringing together these two kinds of measurement—of values and of connotative meanings—and determining what relations, if any, exist among them. The study can also be conceived as a semantic differential analysis of Morris' "ways to live," i.e., of the connotations that the various Ways (as concepts) have for a sample of American college students. Another purpose was to provide evidence for concept-scale interaction (cf. Osgood, Suci, & Tannenbaum, 1957, pp. 176–188); factor analyses of single concepts have indicated that both interscale correlations and even factor structure, to a lesser extent, can vary as a function of the particular concepts being judged, and we wished to see if the same phenomenon would hold for a restricted *class* of concepts—here, a set of value statements. Another methodological interest lay in the complexity of the "concept" being differentiated. Most studies with the semantic differential have required the subjects to react to single words (e.g., hospital, tornado) or at most short phrases (e.g., my doctor, use of the atomic bomb), although complex visual stimuli such as modern paintings have been used; here subjects reacted to very complex verbal statements, averaging 100 or more words, as if to unitary "concepts."

METHOD

Materials

Any form of the semantic differential involves a set of "concepts" and a set of "scales" against which to judge them. In the present case the concepts are a set of 13 statements, each running about 100 words, describing different value-orientations. These statements are the same as those used in other studies of

values by Morris (cf. references above). The semantic scales used in the present study were taken in the main from the 76 used in the Thesaurus factor analysis (cf. Osgood, Suci, & Tannenbaum, 1957, pp. 47–66) to represent the general factors derived in that study: "evaluation" (timely-untimely, kind-cruel, successful-unsuccessful, good-bad, positive-negative, beautiful-ugly, important-unimportant); "potency" (hard-soft, masculine-feminine, strong-weak, powerful-powerless); "activity" (active-passive, quick-slow, warm-cool, excitable-calm); "stability" (true-false, stable-changeable, wise-foolish, predictable-unpredictable); "receptivity" (interesting-uninteresting, savory-tasteless, colorful-colorless); "novelty" (new-old, unusual-usual); and "tautness" (angular-rounded, straight-curved). The actual order and direction, left to right, of these scales is shown in Table 2, this arrangement being maintained for all concepts and all subjects. A separate semantic-differential page was provided for each concept, e.g.,

WAY 1

interesting __:__:__:__:__:__:__ uninteresting
untimely __:__:__:__:__:__:__ timely
new __:__:__:__:__:__:__ old
etc.

and the Ways were judged in the 1 to 13 order given in Table 1.

Subjects

The subjects used in this study were 55 male college students at Stanford University. They were all 20 or 21 years of age. This is a rather homogeneous subsample from the American population, and any generalizations from the present data must be tempered by this fact. The subjects completed the experiment in one session.

Procedure

The written instructions adapted the usual semantic differential instructions for use with the more complex stimuli, paragraph descriptions of the 13 "ways to live." (See Osgood, Suci, & Tannenbaum, 1957, pp. 82–84, for an example of the instructions generally given for the semantic differential.) Following these instructions, the subject found 13 single-page semantic-differential forms, one for each Way. Upon completing his differentiations, he read the following:

> Turn back to the thirteen ways to live which various persons at various times have advocated and followed [each subject had a set of the paragraph descriptions of the Ways].
>
> Indicate by numbers, which you are to write *in the margin*, how much *you yourself like or dislike each of them. Do them in order. Do not read ahead.*
>
> Remember that it is not a question of which kind of life you now lead, or the kind of life you think it prudent to live in our society, or the kind of life you think good for other persons, *but simply the kind of life you personally would like to live.*

Subjects indicated their preferences by writing a number from 7 (I like it very much) to 1 (I dislike it very much) beside each Way.

RESULTS

Preference for and Connotations of the Thirteen Ways of Life

Table 2 gives the mean judgments of each Way against each semantic scale, as well as the mean preference ratings (bottom row), for the 55 Stanford college men. Looking first at the preferences, we see that Ways 7 (integration of diversity), 1 (preserve the best in society), 6 (progress through realistic solution of problems), and 12 (dynamic physical interaction with the environment)—in that order—are the most preferred value-orientations, while Ways 13 (humble obedience to cosmic purposes), 9 (quiet receptivity to experience), and 11 (contemplation of the rich inner life)—in that order—are the least preferred. There are approximately three scale units on the seven-step preference scale separating these most and least liked values, so the preferential discrimination is marked. Stanford men clearly favor diversity, progress, and action in contrast to the more humble, contemplative, and receptive modes of life. That this ordering of value preferences is not unique to our Stanford group is shown by the rank order correlation of their mean preferences with those of Morris' (Morris & Jones, 1955) much larger college student group—a correlation of +.93. The only rank difference of any magnitude is for Way 4 (abandonment to sensuous enjoyment), which is ranked ninth by the larger group and sixth by the Stanford students.

Comparing the mean semantic profiles for the Ways in Table 2 with their verbal statement in Table 1 (and see Morris, 1956b, pp. 15–19) provides rather impressive evidence for the face validity of the judgments on which the profiles are based. Those means that are at least one scale unit away from the midpoints of the scales (4.00), and hence presumably significant deviations from neutral (cf. Osgood, Suci, & Tannenbaum, 1957, pp. 126–140, where significance levels for group data are discussed), are indicated by asterisks in Table 2. Considering these means, we find interesting semantic characterizations for each of the Ways. For example, Way 1 (preserve the best in society) is characterized as old, kind, good, stable, important, calm,

TABLE 2
MEAN VALUES FOR THE THIRTEEN "WAYS TO LIVE"

Scales	Way 1	Way 2	Way 3	Way 4	Way 5	Way 6	Way 7	Way 8	Way 9	Way 10	Way 11	Way 12	Way 13	Opposites
interesting	4.91[a,b]	3.67	4.45	5.31[a]	4.85	5.76[a]	5.55[a]	4.58[a]	3.20	4.64	3.22	5.44[a]	3.05	uninteresting
untimely	3.29	5.04[a]	4.11	3.95	3.78	2.33[a]	2.85[a]	4.07	5.20[a]	3.84	5.15[a]	3.07	5.07[a]	timely
new	2.44[a]	2.65[a]	2.76[a]	3.47	3.75	4.87	4.18	2.95[a]	2.40[a]	3.13	2.64[a]	4.35	2.78[a]	old
kind	5.33[a]	3.15	6.31[a]	3.65	4.24	3.98	4.75	5.13[a]	4.58	3.71	4.36	3.85	4.58	cruel
hard	3.67	5.55[a]	2.51[a]	4.27	4.02	5.07[a]	4.11	2.45[a]	2.60[a]	5.20[a]	3.58	5.04[a]	2.69[a]	soft
unsuccessful	3.04	4.87	4.29	4.16	3.51	2.22[a]	2.76[a]	4.42	5.64[a]	3.75	5.20[a]	2.95[a]	5.44[a]	successful
active	3.96	2.76[a]	3.00	5.25[a]	5.51[a]	6.15[a]	5.04[a]	2.33[a]	1.65[a]	4.11	2.22[a]	6.31[a]	1.89[a]	passive
true	4.82	3.31	4.36	3.82	4.27	5.44[a]	5.02[a]	3.82	2.96[a]	4.24	3.11	4.45	2.76[a]	false
tasteless	3.93	5.18[a]	3.93	3.16	3.67	3.04	2.91[a]	4.02	5.05[a]	4.15	5.09[a]	2.78[a]	5.44[a]	savory
good	5.22[a]	3.27	5.49[a]	3.85	4.42	5.16[a]	5.07[a]	4.33	3.58	4.51	3.75	4.64	3.40	bad
feminine	4.25	3.58	5.04[a]	3.35	3.75	2.60[a]	3.78	4.75	4.84	3.31	4.65	2.04[a]	5.27[a]	masculine
slow	4.82	4.80	4.93	2.91[a]	3.16	2.60[a]	3.53	5.64[a]	5.98[a]	4.07	5.58[a]	2.04[a]	5.38[a]	fast
stable	5.96[a]	4.29	4.84	2.47[a]	3.67	3.47	3.95	5.38[a]	4.93	4.98	4.58	2.73[a]	4.51	changeable
angular	3.58	4.75	3.64	4.49	3.71	4.27	3.53	3.20	3.95	4.18	4.02	4.49	4.02	rounded
cool	4.47	5.20[a]	2.62[a]	3.92	3.51	4.27	3.73	2.78[a]	4.36	4.65	4.53	3.45	3.82	warm
curved	3.36	3.44	3.76	4.53	3.85	3.33	4.24	4.02	4.18	3.53	3.87	3.65	4.31	straight
colorful	3.31	2.71[a]	3.85	5.67[a]	5.07[a]	4.89	4.98	3.55	2.56[a]	3.27	2.47[a]	5.84[a]	2.38[a]	colorless
negative	3.24	5.31[a]	3.07	4.02	3.27	2.53[a]	3.02	4.38	5.11[a]	3.85	4.95	2.75[a]	5.40[a]	positive
unusual	3.33	5.33[a]	4.67	4.16	3.33	3.53	3.60	3.69	4.75	4.44	4.69	4.05	4.45	usual
ugly	3.33	4.65	2.65[a]	3.78	3.78	3.58	3.16	3.25	3.91	4.16	4.04	3.58	4.20	beautiful
strong	4.40	4.07	3.84	3.98	4.35	5.69[a]	4.73	2.89[a]	2.36[a]	5.11[a]	2.75[a]	5.51[a]	2.13[a]	weak
important	5.07[a]	4.20	4.96	4.15	4.55	5.64[a]	5.07[a]	3.65	3.11	4.78	3.18	4.84	2.98[a]	unimportant
excitable	2.40[a]	3.20	2.76[a]	5.31[a]	4.85	4.67	4.20	2.56[a]	2.05[a]	3.09	2.40[a]	5.44[a]	2.44[a]	calm
foolish	2.78[a]	4.98	3.42	4.76	4.18	2.64[a]	3.04	4.22	5.00[a]	3.47	4.91	3.82	5.42[a]	wise
unpredictable	2.07[a]	3.56	3.22	4.96	3.93	3.82	3.98	2.69[a]	3.45	3.18	3.67	4.80	3.65	predictable
powerless	3.36	4.27	3.96	4.07	3.24	2.20[a]	3.33	4.96	5.56[a]	3.16	5.31[a]	2.89[a]	5.96[a]	powerful
Preference: like	5.11	3.02	4.18	4.29	4.06	5.07	5.16	4.36	2.51	4.16	2.66	5.04	2.06	dislike

[a] These means deviate from the scale midpoint (4.00) by at least one scale unit and determine semantic characterizations for each of the Ways (see text).

[b] For values >4.00, the mean judgment is toward the left-hand term; for values <4.00, the mean judgment is toward the right-hand term.

wise, and predictable; Way 5 (group action toward common goals) is active and colorful; and Way 11 (contemplation of the rich inner life) is untimely, old, unsuccessful, passive, tasteless, slow, colorless, weak, calm, and powerless. The connotations of the other Ways can be determined from Table 2 by the interested reader.

Factor Structure of Value Connotations

Since we have each of the 13 Ways judged against the same 26-scale semantic differential, we can determine the structuring of the Ways in terms of their connotative meanings. Two separate analyses were made, one in which intersubject variance was excluded (by using the mean judgments on scales, $N = 26$, as the basis for correlations) and one in which intersubject variance was included (by correlating over both individual subjects and scales, $N = 1430$). In both cases 13/13 correlation matrices, each Way versus every other Way, were computed; these correlation matrices were factored by the centroid method, again using 1.00 communalities, and then were rotated toward simple structure by the quartimax method.[2]

Table 3 gives the quartimax rotation of the first four factors for both types of analysis, based on means (m) and on individual data (i); corresponding factors in the two analyses are placed in adjoining columns for ready comparison. The correspondence of the factor loadings obtained by these two methods is extremely high, as can be seen from inspection

[2] To conserve space, only the rotated factor loading tables are given here. The original correlation matrices and the unrotated centroid loading tables have been filed with the American Documentation Institute. Order Document No. 6692 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

of the paired columns in Table 3. Looking at the absolute size of the loadings, and at the percentage of total variance extracted (bottom row), it can be seen that the analysis based on means extracts more of the total variance on the first two factors while the analysis that includes the intersubject variance spreads the proportion of variance extracted more evenly over the factors, actually extracting more on the fourth factor. These four factors account for 95% of the total variance in the factor analysis of means and for 61% of the total variance in the analysis based on individual data. It is clear that in this case elimination of intersubject variance by use of group means serves to amplify the common variance without seriously distorting the nature of the underlying factors.

What are the common factors in the Ways as revealed through their connotative meanings? Here we infer the nature of a factor from inspection of those Ways which have high versus low loadings on it, and we then check this inference by noting which descriptive scales order the Ways as does the factor. Factor I most clearly divides Ways 9, 11, and 13 (quiet receptivity to experience, contemplation of the rich inner life, obedience to cosmic purposes) from Ways 5, 6, 7, and 12 (group action toward common goals, progress through solution of problems, integration of diversity, dynamic physical interaction); the semantic scales which order the Ways much like this factor are uninteresting-interesting, passive-active, colorless-colorful, and negative-positive. We identify this factor as "dynamism" (the negative pole being dynamic and the positive pole passive). Factor II has Ways 2 and 4 toward the positive end (self-sufficiency, reflection, and meditation; abandonment to sensuous enjoyment) and Ways 1, 3, and 8 toward the negative end (preserve the best in society, altruistic affection and concern, wholesome enjoyment of simple comforts); the most nearly parallel scales are cruel-kind, hard-soft, and angular-rounded. We identify this factor as "socialization" (again reversing the polarity, so that the negative pole is highly socialized and the positive pole highly egocentric).

Factor III segregates Ways 1, 2, and 10 (preserve the best in society; self-sufficiency, reflection, and meditation; dignified self-control) from Ways 4, 8, and 13 (abandonment to sensuous enjoyment, wholesome enjoyment of simple comforts, humble obedience to cosmic purposes); the related descriptive scales are hard-soft and cool-warm. We identify this as a "control" factor (the positive pole being highly controlled and the negative uncontrolled or "abandoned"). Factor IV shows Ways 4, 5, 7, and 12 as considerably positive (abandonment to sensuous enjoyment, group action toward common goals, integration of diversity, dynamic physical interaction) and Ways 2, 8, and 10 as relatively negative (self-sufficiency, wholesome enjoyment of simple comforts, dignified self-control); parallel scales are active-passive, fast-slow, colorful-colorless, and excitable-calm. We more tentatively identify this factor as "venturousness" (the positive pole being the more adventuresome, and the negative pole the more cautious and colorless). It is the least clear of the factors, and differs in some respects in terms of means and individual data.

We are now in a position to characterize these 13 "ways to live" in terms of their locations within this four-factor connotative meaning space, and we can check these factorial characterizations against the verbal statements of the Ways. In other words, we utilize the interpretive economy provided by the factor analysis to describe the 13 value-orientations in terms of four bipolar factors. Table 3 gives these characterizations in the last column. Thus Way 1 (preserve the best in society) is characterized as "controlled socialization" because it has its highest loading on the "socialization" factor and subsidiary loading on the "control" factor. Way 2 (self-sufficiency, reflection, and meditation) is characterized as "egocentric controlled passivity" because it shows loadings on Factors I, II, and III, the highest being on the first. Way 8 is "easy passive socialization" because its highest loading is on the "socialization" factor, its next on "passivity" (nondynamism), and it shows residual loading on "noncontrol" (hence "easy"). Way 13 (humble obedience to cosmic purposes) is characterized as "pure passivity (somewhat abandoned)" because of

TABLE 3

QUARTIMAX ROTATIONS OF THE FIRST FOUR FACTORS FOR CONNOTATIVE MEANINGS OF THE WAYS: ANALYSIS BASED ON MEANS (*m*) VERSUS INDIVIDUAL DATA (*i*)

Way	Factor I (Dynamism)		Factor II (Socialization)		Factor III (Control)		Factor IV (Venturousness)		Factorial Characterization of Ways
	m	*i*	*m*	*i*	*m*	*i*	*m*	*i*	
1	−.09	.12	−.82	−.52	.48	.44	−.16	.22	"controlled socialization"
2	.68	.50	.30	.37	.63	.43	.13	−.08	"egocentric controlled passivity"
3	.17	.20	−.93	−.77	−.01	.08	.10	−.04	"pure socialization"
4	−.54	−.04	.35	.35	−.26	−.34	.69	.56	"abandoned ego-gratifying dynamic venturousness"
5	−.87	−.26	−.14	−.20	−.12	.03	.31	.57	"socialized venturous dynamism"
6	−.95	−.63	−.07	−.03	.27	.42	−.04	.41	"controlled dynamism"
7	−.86	−.32	−.44	−.25	.01	.09	.04	.54	"venturous dynamism"
8	.60	.35	−.74	−.68	−.15	−.09	−.01	−.08	"easy passive socialization"
9	.96	.82	−.23	−.17	.01	−.03	.09	−.01	"somewhat socialized passivity"
10	−.17	−.07	−.21	−.07	.94	.82	−.11	−.07	"somewhat socialized control"
11	.97	.83	−.13	−.06	.12	.14	.09	−.01	"pure passivity (slightly controlled)"
12	−.92	−.51	.17	.05	.06	.07	.27	.58	"egocentric venturous dynamism"
13	.97	.81	−.13	−.16	−.12	−.05	.10	−.04	"pure passivity (slightly abandoned)"
Percentage of Total Variance	55.7	25.2	20.5	13.5	13.2	10.7	5.7	11.5	

extremely high loading on Factor I, yet with a subsidiary loading relatively on Factor III. The reader can check the other characterizations in Table 3 against the factor loadings and against the descriptions of the Ways.

The Semantic Space within Which the Ways Are Judged

The set of 13 value statements used as "concepts" in this study are interesting from a methodological point of view. In the first place, they are a restricted sample of concepts—all of the same class and all highly evaluative ("oughts"). The scales selected for this study provide ample opportunity for at least the three general factors usually obtained, "evaluation," "potency," and "activity," to appear. But do the scale factors show the effects of interaction with the particular set of concepts being judged and differ from those typically obtained with wider samples of concepts? In the second place, these "concepts" are extremely complicated verbal constructions—running 100 words or more in length and obviously shifting in semantic content from phrase to phrase. Do subjects respond to quite different facets of these messages, and hence tend to cancel each other out toward meaninglessness (and an unclear factor structure), or do they distill from each complex the connotative redundancy, and hence reinforce each other toward consistent meaningfulness (and a clear factor structure)?

To get at the factor structure of the semantic scales, we begin by correlating each scale with every other scale across the 13 concepts, generating a 26/26 matrix of intercorrelations. Again, as was the case with the concept factorizations, we make two separate analyses, one based on the means of subject judgments given in Table 2 (where *N* for the correlations is 13) and the other based on summations across both individual subjects and concepts (where *N* for the correlations is 715). These correlation matrices are then factored, using Thurstone's centroid method with communalities fixed at 1.00, and rotated toward simple

TABLE 4

Quartimax Rotations of the First Four Factors for Semantic Scale Relations: Analysis Based on Means (*m*) versus Individual Data (*i*)

Scales	Factor I (Successfulness)		Factor II (Predictability)		Factor III (Kindness)		Factor IV (?)	
	m	*i*	*m*	*i*	*m*	*i*	*m*	*i*
interesting-uninteresting	−.96	−.76	−.15	−.10	−.08	.02	.11	.18
untimely-timely	.97	.79	−.02	.02	.10	.00	−.18	−.22
new-old	−.81	−.48	−.43	−.40	.13	−.01	.11	−.18
kind-cruel	.06	−.25	.34	.30	−.92	−.74	−.10	.00
hard-soft	−.54	−.36	−.02	−.21	.82	.68	−.11	−.10
unsuccessful-successful	.98	.85	−.08	−.03	.00	−.07	−.14	−.15
active-passive	−.91	−.72	−.32	−.47	.18	.20	.00	.11
true-false	−.96	−.85	.20	.09	−.14	−.10	.07	.10
tasteless-savory	.92	.75	.30	.27	.18	.16	−.06	−.13
good-bad	−.79	−.79	.33	.22	−.49	−.30	−.10	−.05
feminine-masculine	.76	.57	.25	.34	−.56	−.49	.07	.03
slow-quick	.84	.59	.45	.59	−.26	−.21	.09	−.05
stable-changeable	.40	.01	.84	.76	−.30	−.18	.19	−.06
angular-rounded	−.02	.16	−.30	−.12	.83	.48	−.40	−.37
cool-warm	.18	.32	.31	.12	.85	.64	.14	.01
curved-straight	−.43	.22	−.70	−.15	−.35	−.28	.27	.46
colorful-colorless	−.80	−.68	−.56	−.48	−.09	−.16	−.04	−.03
negative-positive	.95	.81	−.02	−.02	.27	.22	.10	.06
unusual-usual	.66	.11	.03	.07	.36	.11	−.56	−.45
ugly-beautiful	.46	.53	−.09	−.07	.85	.62	.08	.06
strong-weak	−.93	−.85	.08	−.08	.31	.19	−.14	−.22
important-unimportant	−.93	−.83	.25	.09	.01	−.03	−.17	−.14
excitable-calm	−.71	−.40	−.64	−.71	.23	.18	−.08	.08
foolish-wise	.85	.84	−.44	−.25	.21	.17	−.06	.06
unpredictable-predictable	−.20	.04	−.89	−.68	.29	.03	−.20	−.18
powerless-powerful	.96	.84	−.11	.05	−.18	−.17	.10	.16
Percentage of Total Variance	56.8	38.4	15.8	11.7	19.7	10.8	3.3	3.4

structure, using the quartimax method.[3] Again, as in the factorization of concept matrices above, it was immediately apparent from inspection that the correlations obtained from mean judgments were significantly higher than those obtained with intersubject variance included in the analysis.

Table 4 presents the loadings on the first four rotated factors, the analysis based on means (*m*) and that based on individual data (*i*) appearing in paired columns. Inspection of the figures for percentage of total variance, in the bottom row of this table, confirms the impression from the correlation matrices that the analysis based on means sharpens the common or shared factor structure and accounts for a larger portion of the total variance (by virtue of eliminating the intersubject differences). Only three factors account for 92% of the total variance in the case of means; three factors account for 61% of the total variance in the case of individual data—which is still sizable, of course. Inspection of the values in the paired columns indicates that the two analyses yield exactly the same pattern of loadings. Hence it is again true that no change in the nature of factors is introduced by changing methods of analysis.

The semantic space of connotative meanings generated when these value statements (the Ways) are judged is clearly *not* the same as that obtained when more varied samples of concepts are used. Although Factor I is clearly "evaluative" (timely, successful, true, good, positive, and wise versus untimely, unsuccessful, false, bad, negative, and foolish),

[3] Only the rotated factor loading tables are given here to conserve space. The original correlation matrices and the unrotated centroid loading tables have been filed with the American Documentation Institute. Order Document No. 6692 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

it also includes what are usually independent dimensions—"potency" (strong-weak and powerful-powerless), "activity" (active-passive, quick-slow), and "receptivity" (savory-tasteless, interesting-uninteresting, colorful-colorless). In other words, when different "ways to live" are being judged by American college students, good, strong, active, and stimulating (as opposed to bad, weak, passive, and dull) form a single, overwhelmingly dominant factor in terms of which such value-orientations are discriminated. This prime factor is probably best defined by the single scale, successful-unsuccessful (with loadings of .98 and .85, respectively, in the two analyses). Comparing these results with those of other factor analyses (cf. Osgood, Suci, & Tannenbaum, 1957, Ch. 2), then, we have clear evidence for concept-scale interaction. In the case of value statements (the Ways) being judged as concepts, by American students, "evaluation," "potency," "activity," and "receptivity" fuse together as a single "successfulness" factor.

Two other factors are clearly defined in this analysis, even though their weights are about ½ that of the first factor. Factor II can be identified as a "predictability" dimension from the nature of the scales that load most highly on it: predictable, stable, calm, colorless, slow, old, straight, and wise versus unpredictable, changeable, excitable, colorful, quick, new, curved, and foolish. Factor III appears from the scales that load most highly on it to be a "kindness" dimension: kind, soft, warm, rounded, beautiful, and feminine versus cruel, hard, cold, angular, ugly, and masculine. Factor IV accounts for very little of the variance (3% in both analyses) and is not readily interpretable. Since this was an orthogonal analysis, and since the first three factors together account for 92% (mean data) or 61% (individual data) of the total variance, we may conclude that the semantic space used to characterize the connotations of these 13 "ways of life" is structured along three independent dimensions by American college students—"successfulness," "predictability," and "kindness." It will also be noted that these scale factors are similar in meaningful content to the first three factors derived from the Ways, which were "dynamism," "control," and "socialization."

Prediction of Value Preference from Value Connotation

In this study value preferences and value connotative meanings have been measured independently, in the sense that subjects indicated their preferences by numbers from 7 ("like very much") to 1 ("dislike very much") on the margins of the "Ways" sheets after they had filled out semantic differentials. To what extent can one predict value preferences from the clearly identified connotative meaning factors?

To quantify the relation we must project the preferences and connotative meanings of the 13 Ways in to the same space, and this is possible because the same subjects made both types of judgments. We use the semantic space determined by interscale correlations, which, it will be recalled, was quite exhaustively described by three factors—"successfulness," "predictability," and "kindness." To locate the Ways in this space, we compute the factor scores for each; these are given in Table 5. It can be seen, for example, that Way 1 scores slightly negative on Factor I (slightly successful), considerably positive on Factor II (quite predictable), and slightly negative on Factor III (slightly kind). The first column in Table 5 gives the mean preference ratings of the Ways (taken from the bottom row of Table 2). To the extent that connotations predict preferences, the factor scores (taken either singly or multiply) should correlate with the preference judgments.

These correlations were computed both for the mean preference ratings ($N = 13$) and for the preference ratings given by individual subjects ($N = 55 \times 13$, or 715). The single and multiple r's obtained in both cases are given in the lower portion of Table 5. It can be seen that in terms of both means and individual data, the first factor ("successfulness") correlates highly ($-.93$ and $-.66$) with expressed preferences (Note: the negative correlation simply reflects the way the factor scores are ordered; the actual relation is such that high "successfulness" connotation is associated with high preference). The other two factors display practically no relation to

preference (.09 and −.05 with "predictability" and −.18 and −.04 with "kindness"). As might be expected in this case, the multiple *r*'s (.96 and .66) indicate that the addition of Factors II and III adds little or nothing to the prediction of preferences. We can conclude that the dominant first factor of the semantic space within which the value-statements (the Ways) are discriminated by the Stanford students is the one that predicts value preferences, and that therefore it is, indeed, an evaluative factor.

DISCUSSION

Preferences among various "ways to live" for American college students prove to be very stable *in the cultural sense.* This was shown by the fact that the mean preference ratings obtained here correspond almost perfectly with those obtained earlier by Morris for other groups of American college students. However, in Morris' earlier studies the correlations between Ways, taken across individual subjects, were uniformly low and the factor structures accounted for only a small part (approximately 34%) of the total variance. This is one respect in which judgment of the Ways against the semantic differential has provided clarification: it has accounted for almost all of the total variance and it has indicated that preferences are strongly related to an evaluative factor, which for American college students generally corresponds to the amount of "successfulness" attributed to the alternative Ways. This is especially true when means are used ($r = .93$). When individual data are used, the r of .66, while still high, shows that preference and "successfulness" may at times diverge. (For a further discussion of the relation of preference ratings to evaluations see Morris, 1956b, pp. 163–171.)

Another respect in which this semantic analysis provides clarification is the picture one gets of the connotative framework within which various "ways to live" are considered. Whether one looks at the space generated by the intercorrelations among the Ways themselves or at the space generated by the intercorrelations among the semantic scales, essentially similar connotative factors are sharply etched. Statements of "ways to live" are most clearly discriminated in terms of their "dynamism" (Ways analysis) or "successfulness" (scale analysis); it is this connotation that seems to come through most intensely when subjects read the value statements. Quite independently, subjects discriminate among the value statements in terms of how "controlled" (Ways analysis) or "predictable" (scale analysis) the ways of life seem to be; according to these subjects, there can be "dynamic" Ways that are relatively "controlled" (e.g., Way 6, progress through realistic solution of problems) and "dynamic" Ways that are relatively "uncontrolled" or "unpredictable" (e.g., Way 4, abandonment to sensuous enjoyment). Third, and again independently, subjects discriminate among value statements on the basis of their "socialization" (Ways analysis) or "kindness" (scale analysis). These three general attributes account largely for the differences people see in the connotative meanings of the alternative ways to live.

How adequate a sample of conceivable "ways to live" is the set of 13 used in these studies? The factoring of Ways in terms of their connotative meanings sheds some light on this question, and it becomes most evident

TABLE 5

MEAN PREFERENCE RATINGS AND FACTOR SCORES ON SCALE MATRICES FOR THIRTEEN WAYS

Ways	Mean Preferences	Factor Scores of Ways (from analysis in Table 4 based on means)		
		I (Unsuccessful)	II (Predictable)	III (Unkind)
1	5.11	−0.585	1.994	−0.453
2	3.02	0.758	0.664	2.077
3	4.18	−0.103	0.630	−2.030
4	4.29	−0.130	−2.006	0.278
5	4.06	−0.534	−0.804	−0.264
6	5.07	−1.663	0.296	0.671
7	5.16	−0.965	−0.361	−0.590
8	4.36	0.442	0.242	−1.288
9	2.51	1.517	−0.031	−0.116
10	4.16	−0.310	1.175	1.194
11	2.66	1.191	0.029	0.317
12	5.04	−1.215	−1.207	0.419
13	2.06	1.598	−0.623	−0.216

Correlations between Preferences and Factor Scores

Based on Mean Preferences	Based on Individual Preferences
I/P = −.93	I/P = −.66
II/P = .09	II/P = −.05
III/P = −.18	III/P = −.04
I, II, III/P = .96	I, II, III/P = .66

when we construct a three-dimensional model from the data for means given in Table 3. This model, shown in Figure 1, distributes the 13 Ways in the space according to their loadings on the three factors. We can think of each factor as dividing the space into two regions (or hemispheres) established by a hyperplane passing through the origin at right angles to the factor. The three factors slice our spherical space in three different ways, and, with respect to each way, we can ask if the distribution of Ways is equable. The "dynamism" factor divides the space into upper and lower halves, in terms of the model, and it can be seen that there are about as many Ways above as below this hyperplane. The "socialization" factor divides the space into left- and right-hand halves, and we see that only Ways 2 and 4 (self-sufficiency, reflection, and meditation; abandonment to

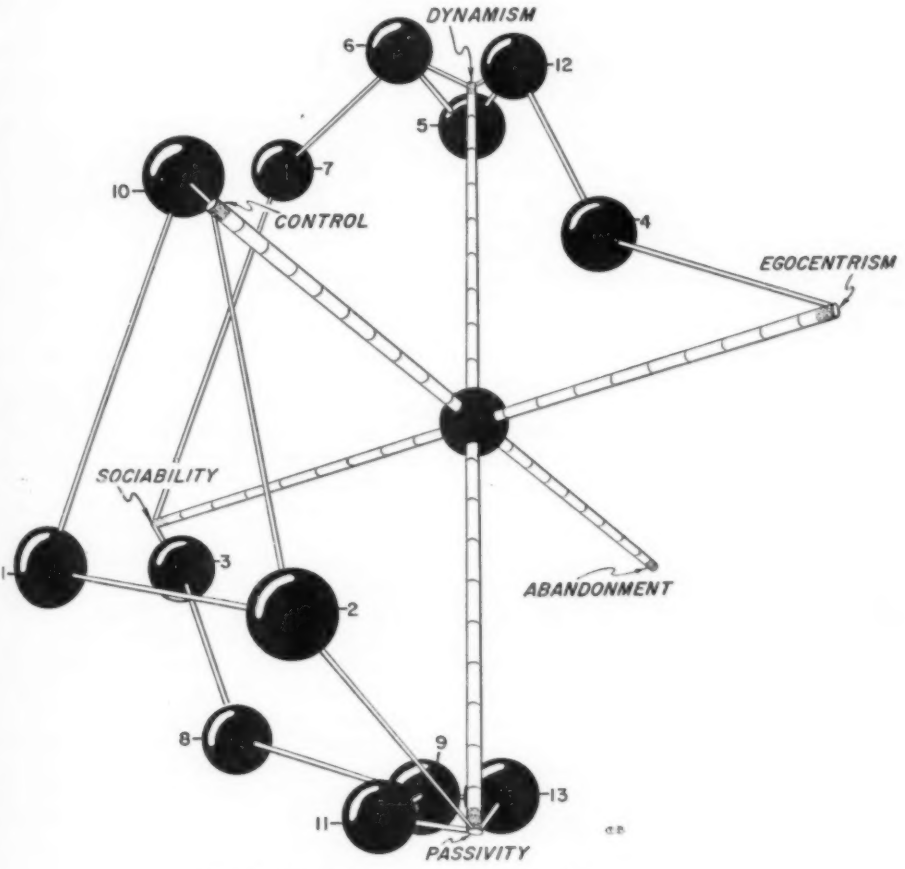

FIG. 1. Three dimensional representation of the ways.

Ways to Live[a]

1. preserve the best
2. self-sufficiency
3. altruism
4. sensuous abandonment
5. group action
6. progress through realism
7. integration
8. simple comforts
9. quiet receptivity
10. self-control
11. contemplation
12. physical interaction
13. humble obedience

[a] See Table 1 and *Varieties of Human Value*, pp. 15–19, for more complete description.

sensuous enjoyment) fall in the right-hand, "egocentric" regi The "control" factor divides the space into front and rear halves, as the model is viewed, and it is apparent that only Way 4 (abandonment to sensuous enjoyment), and to a slight extent Way 8 (wholesome enjoyment of simple comforts), fall in the "noncontrol" or "abandonment" region.

It is clear, then, that the connotative meaning space for "ways to live" is not uniformly populated by the Ways treated in this and Morris' earlier studies. Specifically, it is empty of conceptions of "the good life" which stress extremes of "egocentrism" and/or "noncontrol." We can imagine Ways that might fit these criteria—"The only way to live which is both rational and maximally gratifying is *narcissism*; one should live so as to appreciate and glorify the self, since Ego is the only true reality . . ." or "The goal of life should be social, emotional, and intellectual *anarchy*; each individual is an isolated island in fact and should be in spirit, always resisting social conformity, emotional dependence, and intellectual tyranny . . . ." But as soon as we do try to imagine Ways falling in these regions, it becomes apparent why they are not likely to appear in the value-orientations of *organized groups of humans*. The establishment and maintainance of societies almost of necessity implies *some* degree of both socialization and control over the individual. Hence conceptions of "the good life" that characterize societies do not include narcissism or complete anarchy. On the other hand, for the purposes of measuring values, it would be wise to include value statements like the foregoing to "round out" the space. The model also suggests another possible inadequacy. At least for American college students, Ways 5, 6, and 12, on the one hand, and Ways 9, 11, and 13, on the other, are connotatively redundant (undifferentiated) within the system of 26 scales used here.

However, all of what has been said so far in this discussion takes no account of either individual or cultural differences. Magnitudes of correlations, of factor loadings, and of percentages of total variance accounted for by factors were found to decrease sharply when interindividual variance was included in the analysis. Although this, in part, reflects the greater reliability of mean judgments as compared with individual judgments (and thus decrease in the error variance), it also may reflect individual differences in the assessment of values. One of the important directions which research on the connotative meanings of the Ways should take in the future is into individual differences—in demographic status (age, sex, socioeconomic level, etc.), in personality, in temperament, and the like.

The lack of cross-cultural comparative data also places limits on the generality of our findings. Although Morris (1956b, Ch. 3) has found evidence for common factors in value preference, despite differences in the preferential ordering of the Ways, for cultures as disparate as India and the United States, we cannot assume that this will hold for the connotative framework within which such value-statements are interpreted. In fact, it seems unlikely that the dominance of a "successfulness" factor (combining evaluation, potency, activity, and receptivity) would hold for certain other cultures, such as, say, India. And would subjects drawn from other cultures display the same clustering of Ways according to connotative meanings, (even though differently preferred), as shown in Figure 1 for American college students? Would the tight bundles composed of Ways 5, 6, and 12 versus Ways 9, 11, and 13 prove to be more widely discriminated, say, by Moslem Arabs? Comparative cross-cultural study of value preferences and connotative meanings is another direction for future research on values—research that could have significant bearing on international communication and understanding.

Finally, there are three matters of methodological interest that deserve mention. First, the results of the factor analysis of scale relations make it clear that factors derived from more "representative" samples of concepts are not necessarily independent, and hence visible, when some specific subset of concepts is judged. The semantic scales used in this study were selected as representing seven independent factors obtained in a previous analysis based on 100 subjects judging 20 varied concepts against 76 scales; in the present study we find a quite different

factor structure, for the same scales and type of subject. Concept/scale interaction thus holds for groups of related concepts as well as for single concepts, and this may mean that we must devise different semantic differentials for different classes of concepts—value concepts, political concepts, personality concepts, esthetic concepts, and so on. Second, we have found that analyses based on correlations over means and analyses based on correlations over individual subjects yield almost identically the same factors. Although this relationship need not always be the case (indeed, one can arrange hypothetical data in such a way that it would not be true), it does increase the likelihood that comparisons made in the past by Osgood and his collaborators between factor analyses of these two types are feasible. Third, the fact that we have obtained clear-cut factor structures in both Ways and scale analyses, despite the length and complexity of the "concepts" being judged, clearly indicates that complex verbal statements of this sort can be studied effectively with the semantic differential.

SUMMARY

Standard statements of Morris' 13 "ways to live" were judged against a 26-scale form of the semantic differential by 55 Stanford college men. These subjects also indicated their preferences among these "ways to live." The average ranking of *preferences* for the Stanford group was almost identical with that for other American college student samples studied by Morris and his associates. Factor analyses of the *connotative meanings* of these "ways to live," as measured by the semantic differential, yielded a reasonably clear, four-factor system: "dynamism," "control," "socialization," and perhaps "venturousness." Factor analyses of the semantic scale relations yielded three clear factors that accounted for large portions of the total variance: "successfulness," "predictability," and "kindness." Two types of factor analyses, one based on correlations over means and the other on correlations over individual subjects, were made, and for both the concept and the scale factorizations identical factors were obtained. It was found that the preferences for "ways to live" could be predicted from the factor scores on the first connotative factor ("successfulness") to the extent of a correlation of .93 (for average preferences) and .66 (for individual preferences). Substantive and methodological implications of these findings are discussed.

REFERENCES

JONES, L. V., & MORRIS, C. Relations of temperament to the choice of values. *J. abnorm. soc. Psychol.*, 1956, **53**, 345–349.

KLUCKHOHN, C., et al. Values and value-orientations in the theory of action. In T. Parsons & E. A. Shils (Eds.), *Toward a general theory of action*. Cambridge: Harvard Univer. Press, 1951.

KUMATA, H. A factor analytic investigation of the generality of semantic structure across two selected cultures. Unpublished doctoral dissertation, University of Illinois, 1957.

MORRIS, C. *Paths of life*. (Originally published, Harper 1942) New York: George Braziller, 1956. (a)

MORRIS, C. *Varieties of human value*. Chicago: Univer. Chicago Press, 1956. (b)

MORRIS, C., & JONES, L. V. Value scales and dimensions. *J. abnorm. soc. Psychol.*, 1955, **51**, 523–535.

OSGOOD, C. E., & LURIA, Z. A blind analysis of a case of multiple personality using the semantic differential. *J. abnorm. soc. Psychol.*, 1954, **49**, 579–591.

OSGOOD, C. E., & SUCI, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49**, 251–262.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana, Ill.: Univer. Illinois Press, 1957.

OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychol. Rev.*, 1955, **66**, 42–55.

TRIANDIS, H. C., & OSGOOD, C. E. A comparative factorial analysis of semantic structures in monolingual Greek and American college students. *J. abnorm. soc. Psychol.*, 1958, **57**, 187–196.

(Received December 4, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 74-78

REDUCTION OF COGNITIVE DISSONANCE BY SEEKING CONSONANT INFORMATION[1]

J. STACY ADAMS[2]
Stanford University

WHEN a person holding an opinion on a matter of importance is exposed to an authoritative communication advocating a contrary opinion, cognitive dissonance and a need to reduce the dissonance are produced (Festinger, 1957). Several means of dissonance reduction are possible: The person may change his opinion to conform more closely with the communication; he may attempt to persuade the communicator to alter *his* opinion; he may distort the content of the communication; he may discredit the communicator; he may reduce the importance of his opinion; or he may attempt to obtain support for his opinion. It is with the last of these modes of reducing dissonance that we are concerned in this study.

Previous studies have shown this mode in operation. Festinger, Riecken, and Schachter (1956) found that a group of people, in whom strong dissonance had been produced by the disconfirmation of their beliefs, began to recruit adherents to their cause. Their efforts to obtain social support were in marked contrast to their behavior prior to the onset of dissonance, when they were reluctant to take in new members and reticent in releasing news about their beliefs to the press. Maccoby (1959) found evidence of seeking consonant social support in two field experiments conducted with mothers. Her data show that mothers in whom dissonance was greatest, as a result of exposure to a contrary communication on child training, were most likely to have discussed the topic of the communication with like-minded persons. In an experimental study of the effects of group discussion on counteracting propaganda, Brodbeck (1956) showed that persons in whom dissonance had been produced displayed a preference to engage in discussion with nonexpert peers whose beliefs were consonant with theirs, thus obtaining social support and restoring confidence in their beliefs to prepropaganda levels. Because of the experimental design, Brodbeck was not able to show that seeking social support increased as a function of magnitude of dissonance.

In a different, but related, vein, Ehrlich, Guttman, Schonbach, and Mills (1957) found that persons with postdecision dissonance resulting from the purchase of an automobile were highly selective in the car advertisements they noticed and read. Specifically, they were more likely to notice and read ads about their own make of car than about other makes, and were more likely to read ads about their car than were car owners with old cars. In this case, seeking consonant information rather than seeking social support was the approach taken to reduce dissonance. In a further study Mills, Aronson, and Robinson (1959) found dissonance-reducing selectivity to positive information among students who had made a decision to take one of two alternative types of examinations.

From Festinger's theory and the foregoing data the following two hypotheses are derived for testing. First, a person in whom dissonance has been produced by exposure to an authoritative communication espousing a contrary point of view on an important subject, is more likely to seek authoritative information on the subject than a person who has been exposed to a consonant communication. This prediction is expected to hold only if other means of reducing dissonance are relatively less available than seeking informational support. Furthermore, the prediction is expected to hold only if the communication is authoritative and if its subject matter is important to the subject. If the communication were not perceived as at least minimally authoritative (trustworthy, credible, expert), or, similarly, if the subject of the person's opinion were not important to him, little

[1] The research was supported in whole by United States Public Health Service Grant 3M-9122C.

[2] I wish to express my thanks to Leon Festinger for his continuing interest, support, and advice during the study; to A. K. Romney for his assistance in a preliminary study; and to Nathan Maccoby for giving the post-experimental talk.

dissonance should result. The hypothesis receives some, though unsatisfactory, support from the field studies of Festinger et al. (1956) and Maccoby (1959). In the first a control group not experiencing dissonance was lacking. In the second study it was only shown that seeking consonant support varied as a function of dissonance; it was not shown that seeking support per se varied with dissonance.

The second hypothesis concerns the selectivity of information-seeking. A person in whom dissonance has been produced by exposure to a contrary communication seeks support, if such is available, from a source that is perceived to agree with his opinion. If the person's opinion is not substantially changed by the communication, he should seek support from a source perceived to agree with his precommunication opinion. If, on the other hand, his opinion is changed substantially to conform with the communication, he should seek support from a source perceived to agree with his new opinion. This is derived from the fact that, once the person has partly reduced dissonance by adopting the opinion advocated by the communication, new dissonance results, produced largely by the psychological incompatibility of the new opinion with cognitions related to the original opinion. In sum, changing opinion produces new cognitive dissonance that can effectively be reduced by obtaining support for the new opinion. This hypothesis receives direct and indirect support from the data of Ehrlich et al. (1957), Maccoby (1959), Mills et al. (1959), and Brodbeck (1956). However, in the Brodbeck study, which is the most relevant to the present experiment, subjects were constrained to communicate *either* with a subject who held consonant opinions or one who held dissonant opinions, whereas in the present study subjects were free to choose whether they wished information or not, in addition to being free to select the source of information. In addition, the discussants in Brodbeck's study were nonexpert peers, whereas in the experiment reported here the sources available to the subjects were experts.

METHOD

The subjects were 100 mothers whose youngest child was in the first grade of five Palo Alto, California, public schools. Fifty-three of these, randomly chosen, were experimental subjects; 47 were control subjects.[3] Each subject was put through the following experimental procedure in her home by one of four trained interviewers.[4]

1. The subject was asked a series of opinion questions on the relative importance of hereditary and environmental factors in child development, in addition to a few background questions and some opinion questions on the issue of permissiveness in child rearing. The key question for the purpose of the experiment was a 10-point scale that followed eight questions on the issue of heredity-environment:

> On the whole, would you say that a child's behavior is mostly learned or would you say that it is mostly inborn? Show your opinion by putting a cross (X) at either end, or any place in between, on the line below.
>
> Mostly Learned ⊢—⊢—⊢—⊢—⊢—⊢—⊢—⊢—⊢—⊢ Mostly Inborn

Following this was the question, "How sure are you of the answer you gave to the last question?" to which five answers ranging from "very sure" to "not at all sure" could be given.

2. The subject was then asked to listen to an 18-minute tape-recorded talk by a fictitious expert, Dr. Wilkins, emphasizing either hereditary or environmental factors in child development. Both the hereditary and the environmental talk were attributed to the same fictitious expert, for whom a brief biographical sketch was given, and were recorded by the same person. Of approximately the same duration and making about the same number of points, the talks differed only in the point of view advocated—pro-heredity or pro-environment. All factual data referred to had origins in the published literature; no fictitious information or misinformation was given.

Experimental subjects were given whichever of the two talks was contrary to the opinion they expressed on the 10-point scale. For this purpose the subject was considered as having an environmental opinion if she checked any of the five intervals on the left of the scale ("Learned" pole) and as having a hereditary opinion if she checked any of the five intervals on the right of the scale ("Inborn" pole). If the subject placed her check precisely in the middle of the scale, a systematic scheme determined which talk she heard. Control subjects were given whichever of the two talks agreed with their opinion.

3. Following the talk, the subjects were told:

> You have heard Dr. Wilkins' views on "Heredity and Environment." There are, of course, other views that are just as authoritative. Now, we'd like to give you the opportunity to discuss some of the things you heard in the talk, if you want to. Of

[3] A few mothers gave birth to a child between the time the sample was drawn and the study was conducted. Three mothers were mistakenly treated as experimental subjects instead of as control subjects.

[4] Approximately equal numbers of experimental and control subjects were assigned at random to each interviewer. The interviewers were unaware of the specific hypotheses under test.

course, it's up to you. We thought some of you might want to have a discussion, so two people have agreed to talk further on the subject. Both are psychologists with interests relevant to the talk you heard and will be at Stanford in about two weeks."

A brief biographical sketch of each psychologist, emphasizing his expertness, was then read. Except for their names, the speakers were differentiated only by the title of a noted article they had allegedly written. For the pro-heredity speaker the article was entitled, "The Importance of Inheritance and Maturational Factors in the Development of Children"; for the pro-environment speaker it was "The Role of Training and Environmental Factors in Child-Rearing." Both speakers were to give their talks in different rooms, but on the same evening, at the same time, and in the same building.

The subject was then asked: "Since for planning purposes we need to know approximately how many people will attend each of the talks, would you tell me if you want to attend either of the talks and, if so, which one?" The interviewer recorded the subject's preference for either (or neither) speaker and, in addition, recorded all comments made by the subject that would permit an estimation of the degree of the subject's desire to attend the talk of her preference.

4. At the conclusion of the session the subject was asked for her opinion again and her confidence in it. This was done under the guise of the interviewer's checking over the original questionnaire and discovering she had "failed" to record the subject's answers to two questions. At this point the interviewer apologized to the subject and asked her, "Would you please answer these now?"

Approximately 2 weeks following the experimental sessions, a talk on the question of hereditary versus environmental factors in child rearing was actually given for those subjects who availed themselves of the opportunity. At this time the object of the experiment was also explained. Both the talk and the description of the experiment were received with considerable interest. However, as is the case in many field experiments in which there are no constraints, very few subjects attended the talk.

The subjects, taped talks, and experimental procedure described above were well-suited to a test of the hypotheses. Dissonance was assured by a topic for the talk that would be important to the mothers of 6- or 7-year-old children. The taped talks, however, were not so rabidly partisan on either side as to be easily discredited. Nor were they so persuasive as to produce substantial opinion change, thus minimizing opinion change and maximizing information-seeking as modes of dissonance reduction. On the other hand, the talks were explicit enough in the points of view advocated so that distortion of content would be difficult. Attempts on the part of the subjects to reduce dissonance by trying to alter the speaker's opinion were, of course, impossible, and seeking social support from the interviewer was effectively denied the latter's not interacting with the subject.

RESULTS AND DISCUSSION

With the exception of four experimental subjects and two control subjects, all subjects expressed an initial opinion on the environmental ("Learned") side of the scale. Among the 94 subjects indicating an environmental opinion, 34 experimental and 32 control subjects checked one of the three most extreme intervals on the scale. It is with these subjects that the main analysis will proceed, as this maximizes the opportunity of revealing differences in the groups as a function of the experimental manipulation. Since these subjects all had environmental opinions, it follows that experimental subjects heard a hereditary communication and that control subjects heard an environmental communication.

Success of the Experimental Manipulation

Though the communications were selected because they would not be so persuasive as to produce large amounts of opinion change, it might be expected that if the communications were indeed different and productive of dissonance in the experimental group, the latter would show a significantly greater amount of opinion change in the direction of heredity than would the control group. For this purpose the number of subjects in each condition changing opinion in a hereditary direction may be compared. Sixteen experimental subjects as contrasted to five control subjects thus changed. By chi square test the difference is significant at the .01 level.[5] As a further indication of whether or not dissonance was produced among experimental subjects changes in confidence following the communication may be observed. If the hereditary communication was productive of dissonance for experimental subjects, more experimental subjects should show decreases in confidence in their opinion. The data bear this out, as six experimental subjects but only one control subject showed a decrease in confidence. The difference is significant at the .06 level by Fisher's exact test but is difficult to interpret because all subjects but one showing decreases in confidence also changed opinion slightly.

[5] All chi square tests of significance are two-tailed. The null hypothesis is rejected when $p > .05$.

We may, nevertheless, conclude that the experimental manipulation was successful.

Information Seeking

The first hypothesis to be tested stated, briefly, that persons in whom dissonance had been produced by a communication would be more likely to seek information than persons in whom a communication had produced no (or little) dissonance. An appropriate test consists of a comparison of the number of subjects in each condition who expressed some or no desire to attend a talk, irrespective of any consideration of the alleged point of view of the speaker. Each subject was therefore rated by two coders independently as to whether she expressed some or no desire to attend one of the talks. The coders had no knowledge of the hypothesis being tested and agreed on 65 of their 66 ratings. The single disagreement was resolved by the experimenter against the hypothesis. The data in Table 1 show that experimental subjects were more likely than control subjects to want to attend an authoritative talk. The difference between the groups is significant at the .02 level by Fisher's exact test. It is clear, therefore, that under conditions of *free choice* subjects attempt to reduce dissonance by seeking information.

Selection of Information Source

When the subjects were asked if they wished to attend the talk, they indicated which of two speakers they would prefer to hear—one of them by implication holding a hereditary view and the other an environmental view. It was hypothesized that the experimental subjects would elect to hear a talk consonant with their opinion following the communication. Since only one subject changed opinion entirely, going from an environmental to a hereditary opinion, the data will be analyzed only for subjects who initially had extreme environmental opinions, who either maintained their initial opinion or changed opinion by a small amount only, who wished to hear a talk, and expressed a preference for one of the speakers. To test the hypothesis we may compare the observed distribution of preferences of experimental subjects against an equally split distribution which would be expected if no differential preferences existed. This is shown in Table 2. The difference is significant at the .06 level only.

Though a greater preference for a consonant (i.e., environmental) talk among experimental subjects is suggested by the data, is this greater preference a function of the amount of dissonance produced? In answer we may compare the preferences of experimental and control subjects. The data are given in Table 3. No significant difference in the preferences of the two groups exists; both clearly prefer to hear the consonant, environmental talk. This is contrary to the hypothesis and to previous findings (Brodbeck, 1956; Ehrlich et al., 1957; Maccoby, 1959; and Mills et al., 1959), but an explanation is available if we assume that the subjects who had no desire to hear a talk were those who had no, or at least very little, dissonance. Thus, all or most subjects in Table 3 have some dissonance. As a result, we may expect that both groups should tend to avoid exposure to the dissonance-increasing hereditary talk and seek exposure to the consonant, environmental

TABLE 1
DESIRE TO ATTEND TALK

Desire	Experimental	Control
Some desire	33	25
No desire	1	7

$p = .02$, by Fisher's exact test.

TABLE 2
OBSERVED AND HYPOTHETICAL PREFERENCES OF EXPERIMENTAL SUBJECTS FOR POSTCOMMUNICATION TALK

	Environmental Talk	Hereditary Talk
Observed preference	22	8
Equal preference	15	15

$p = .06$ by chi square.

TABLE 3
PREFERENCES FOR POSTCOMMUNICATION TALK BY EXPERIMENTAL AND CONTROL SUBJECTS

	Environmental Talk	Hereditary Talk
Experimental Subjects	22	8
Control Subjects	19	5

talk, and that the preferences of the two groups should not differ greatly. Nevertheless, since the experimentals should have more dissonance on the average than the controls, we would expect them to show a somewhat greater relative liking for the environmental talk. The tendency is actually in the opposite direction, as Table 3 shows.

SUMMARY

An experiment was conducted to test two hypotheses about the reduction of cognitive dissonance by seeking information. The hypotheses were: (*a*) A person in whom dissonance has been produced by exposure to a communication advocating an opinion contrary to the person's is more likely to seek information than a person exposed to a compatible communication, and (*b*) a person in whom dissonance has been produced by a contrary communication tends to seek information from a source agreeing with his opinion.

The opinions of 100 mothers on the importance of hereditary and environmental factors in child rearing were ascertained by personal interview; they were then exposed to a tape recorded, authoritative communication espousing a hereditary or an environmental point of view. Following the communication they were asked if they wished to attend a talk by an expert allegedly favoring an environmental view or one favoring a hereditary view. For approximately half the subjects the tape recorded talk was contrary to their opinion; for the remaining subjects the talk agreed with their opinion.

The results supported the first hypothesis. The second hypothesis was not sustained, though most subjects indicated a preference to hear a talk by a consonant source. A possible explanation for the failure of the data to support the second hypothesis was given.

REFERENCES

BRODBECK, M. The role of small groups in mediating the effects of propaganda. *J. abnorm. soc. Psychol.*, 1956, **52**, 166–170.

EHRLICH, D., GUTTMAN, I., SCHONBACH, P., & MILLS, J. Postdecision exposure to relevant information. *J. abnorm. soc. Psychol.*, 1957, **54**, 98–102.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston: Row, Peterson, 1957.

FESTINGER, L., RIECKEN, H., & SCHACHTER, S. *When prophecy fails.* Minneapolis: Univer. Minnesota Press, 1956.

MACCOBY, ELEANOR. Social reinforcement and attitude change. Unpublished study, Department of Psychology, Stanford University, July 1959.

MILLS, J., ARONSON, E., & ROBINSON, H. Selectivity in exposure to information. *J. abnorm. soc. Psychol.*, 1959, **59**, 250–253.

(Received December 14, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 79-93

FAMILIAL CORRELATES OF AGGRESSION IN NONDELINQUENT MALE CHILDREN

WILLIAM McCORD, JOAN McCORD, AND ALAN HOWARD[1]
Stanford University

CONSERVATIVE social philosophers from Hobbes through Ortega y Gasset have had a profound suspicion of the motives and capacities of the "mob." Their dislike for the mass has been premised on a series of assumptions concerning human nature. One of the most fundamental of these axioms has been the conviction that man is innately aggressive. Hence, in the opinion of these theorists, the instinctive cruelty and barbarism of mankind must be kept under the strict control of a more enlightened "élite"—a class, who, through long training in the ways of civilization, has inhibited its own aggressive impulses.

This jaundiced view of human motivation has been buttressed by the Freudian belief that aggression is a universal outcome of the "thanatos instinct." In Freud's view—at least, in one of its several versions—all men strive not only for libidinal satisfactions, but also for return to a state of nothingness, of "nirvana." Freud argued that aggression, a desire to destroy, was the "natural" energy emerging from the death instinct; a drive which makes its appearance in infancy and is either fused with "Eros," directed toward the destruction of others, or aimed at self-annihilation.

On a general level, this view of aggression as an inevitable element in the make-up of man has been challenged by such writers as Fromm, Horney, Allport, and Maslow. These observers have put forward the position that aggression is the result of specific types of social milieu, rather than being an instinctual drive.

In an empirical fashion, other investigators have produced evidence which—at least, implicitly—contradicts the conservative Freudian opinions about aggression. Many analysts have demonstrated the variability of the level and expression of aggression in response to various environmental conditions (cf. Weiss & Fine, 1956). Others have concentrated on the child rearing antecedents of extremely aggressive antisocial behavior. For example, the Gluecks (1950), examining a lower class sample, and Bandura and Walters (1959), concentrating on a middle class group, have independently produced a similar portrait of the backgrounds of such individuals. Both studies found that the aggressively antisocial person emerged from an environment characterized by parental rejection, familial discord, punitive discipline, and inconsistency.

Perhaps the most comprehensive study of the early determinants of aggression can be found in the work of Sears, Maccoby, and Levin (1957). From carefully executed interviews with 379 New England mothers (and other sources of data) these researchers concluded that aggression in small children—defined as "behavior intended to hurt or injure someone"—was associated with such environmental antecedents as parental permissiveness for aggression, the use of physically punitive discipline, and maternal lack of self-esteem. These results led the investigators to reject the instinctual view of aggression and, instead, to affirm: "Today, with our better understanding of the ruling influence of culture as a source of behavioral universals, we are inclined to suspect that the determination of which stimuli will arouse rage, and the focusing of the emotion into an actual *desire and intent* to hurt others (or the self), are products of learning experiences begun in early infancy" (p. 222).

The purpose of this paper is to report a study which supports the view that aggression is a form of behavior which is learned by the child during his early experiences within the family. This research, concerned with 174 boys and their families, differs from certain of the previous studies: none of the subjects were known delinquents; the sample was drawn from a relatively lower class, urban

[1] The authors wish to express their appreciation to the Ella Lyman Cabot Foundation, the Harvard Laboratory of Social Relations, and the United States National Institute of Mental Health (Grant M-2647) for their generous financial support of this research.

environment; and the information on child rearing practices and on the boy's aggression was gathered through direct observation of both parents and child over more than a 5-year period. Each of these differences is important in assessing the significance of the research:

1. The fact that delinquents were excluded from the sample means that this study of "normal" aggression is not obscured by a simultaneous examination of the causes of delinquency.

2. The fact that the sample is drawn in disproportionate numbers from the lower class has two advantages: the sample serves as counterbalance to those findings which have emerged from analyses of primarily middle class children; and, at the same time, it counterbalances studies of delinquency which have usually concentrated on lower class samples.

3. The fact that the material was gathered through direct observation eliminates the difficulties that hamper studies based on interviews, experimental situations, or therapeutic sessions.

Background of the Research

The opportunity to investigate the genesis of aggression came from the Cambridge-Somerville Youth Study of the 1930s. This extensive project, aimed primarily at the prevention of delinquency, has been described in numerous other publications (McCord, McCord, & Gudeman, 1960; McCord, McCord, & Zola, 1959; Powers & Witmer, 1950); consequently, we will outline its history only briefly.

Six-hundred and fifty boys, referred by their teachers, ministers, or by social agencies, were chosen for participation in the experiment. At the time of selection, the boys averaged 9 years in age and were drawn primarily from lower class districts in Cambridge and Somerville, Massachusetts. In composition, the sample contained a relatively high proportion of Catholics and recent immigrants, approximately 25% of the families were on relief during the depression, and only 4% of the parents had attended college. It should be noted, in passing, that these "peculiarities" of the sample did not generally affect the results. Neither social class, immigration patterns, ethnic group, nor paternal occupation was significantly related to the child's aggression.

Since the subjects were selected from a restricted socioeconomic level, the lack of association between the child's behavior and more general social factors is hardly surprising.

Half of the boys were selected because they were judged potentially delinquent, "maladjusted" children; the remaining cases were regarded as "normal," nondelinquent boys.

Between 1939 and 1945, 325 children (composed equally of delinquent and normal boys) were given medical, educational, and social aid; all of these boys were assigned to a counselor who visited and advised both the boys and their families. The remaining 325 cases, closely matched to those who received treatment, formed a control group.

In the process of the study, the boys were subjected to physical, psychiatric, psychological, and intelligence examinations. Reports on the boys and their families were systematically gathered from social agencies, the schools, and various community agencies. Detailed reports on the behavior of the boy and of his family were recorded by the staff after each contact.

In 1956, the authors and their colleagues initiated a re-evaluation of the Cambridge-Somerville project, aimed at the solution of three problems:

1. What effect did the treatment have on the adult behavior of the subjects?

2. What correlations could be detected between the early familial experiences of the boys and their later behavior as adults?

3. What relationships could be noted between the familial environments and the boys' behavior in childhood?

As a first step toward resolving these issues, the cumulative case histories of the subjects were read by a trained staff (who did not have knowledge as to the adult behavior of the boys). A rating system concerned with 150 aspects of the boys' families, physical status, general social condition, and personality was developed. Each category as far as possible was defined in terms of observable behavior.

The Measure of Aggression

"Aggression" is an amorphous term. It has been used to describe such seemingly disparate phenomena as gossip, jokes, suicidal tendencies, hostile fantasies, as well as directly destructive acts. A simple event—slapping a person on the back—can be viewed by one observer as a hostile attack, while another individual, aware of the context surrounding the action, would regard it as evidence of exuberant friendship. Thus, the first task in any study of aggression is to define the concept in behavioral terms which can be replicated by other investigators. For the purposes of this research, we followed Sears, Maccoby, and Levin in defining aggression as "behavior intended to hurt or injure someone." The word "intention" is, itself, tricky; it is notoriously difficult to establish the underlying desires which prompt an action. Consequently, we sought to limit the term aggression to those acts which, objectively, "hurt or injured someone."

Fortunately, the staff of the Cambridge-Somerville study had ample opportunity to observe the boys' behavior. Counselors visited their homes and neighborhoods and watched their behavior in relation to parents,

friends, neighbors, and siblings. The psychiatrists, physicians, and psychologists submitted reports on the boys' reactions under relatively stressful conditions. The boys' teachers and other school officials were systematically questioned concerning the children's activities. Camp counselors, social agencies, ministers, police, YMCA directors, and employers submitted additional information. Not only did the information come from diverse sources, but it was gathered over an average period of 5¾ years—thus making it possible to observe the boy under different conditions, at different ages.

From these various sources of information, the 174 boys were placed into one of three categories.[2]

Twenty-five boys were consistently and overtly *aggressive*. These individuals participated in the whole range of aggressive acts; they were involved in fist-fighting, bullying of smaller children, attacks on their teachers, or destructive actions in camp or in the community. They frequently threatened other children with violence—and often carried out their threats. In response to almost any form of frustration, they reacted with verbal abuse, open rage, and attempts to destroy the frustrating object.

Ninety-seven boys were, by the norms of American society, normally *assertive*. They participated, at times, in fights, acts of destruction aimed at other children or adults, occasional battles with teachers or other community officials, and bullying of weaker children or girls. They differed from the openly aggressive boys, however, in that their hostile responses were sporadic exceptions to the general pattern of their lives. Normally, in the face of frustration, they responded realistically; sometimes with anger, but seldom with aggressive actions. In the classroom, they sometimes created disciplinary problems for the teacher, but usually they conformed to the standards of the school.

Fifty-two boys were consistently *nonaggressive*. Outbursts of rage and directly aggressive attacks were exceptionally rare in their lives; and, indeed, in some cases, nonexistent. In the classroom, they did not actively participate in aggressive disturbances. In relations with other children, they were placid and friendly. They could absorb the usual frustrations of childhood—failure in school, loss of a game, discipline from an adult—with apparent calm and realism. Thus, this categorization of aggression is descriptive not only of isolated acts, but of the entire *pattern* of the child's behavior over a period of 5 years.

Two raters, independently categorizing the behavior of 30 cases randomly selected from the sample, reached 86.7% agreement in their judgments of aggression.

The Measures of Family Environment

Our task was to assess the relevance of the child's early environment in determining whether he became "aggressive," "assertive," or "nonaggressive." To achieve this goal, the "raw" information gathered by the staff in their 5 years of direct observation of the families was systematically categorized into a set of variables, descriptive of the child's environment. These ratings ranged from such relatively objective matters as the occupation of the parents, their education, and ethnic group, through more subjective judgments concerning emotional relationships within the family. Categories were developed concerning the parents' attitudes toward one another, their role in the family, their affective relationship with the son, and their personality. The parental methods and consistency of discipline were considered, as well as the specific values which they wished to inculcate. The child's physical status and aspects of his behavior, other than aggression, were also evaluated. (It should be noted that none of the measures of the child's physical status—the conditions of his birth, neurological or glandular disorder, early childhood diseases, body deformities, etc.—was significantly associated with the level of aggression.)

Although the specific definitions and reliability of each category will be discussed in the following pages, certain general characteristics of the ratings should be mentioned:

1. The observations were made in the home of the child, over an average period of 5¾ years, by a number of different people (usually, each child was exposed to at least three counselors). These observers did not know that their material would be used for a study of aggression; hence, the possibility of conscious bias or of a "halo effect" was minimized.[3]

[2] From the original sample of 650, 476 were eliminated for the following reasons:

70 died, moved from Massachusetts, or were dropped at an early stage from the treatment program. The 70 matched mates of those who were eliminated from treatment group were omitted from the control group.

255 boys formed the remainder of the control group. Because information on them was less extensive (and, we believe, less valid), they were dropped from this research.

74 boys were known delinquents who had been convicted of serious crimes. These boys were eliminated for we did not wish the analysis of aggression to be obscured by a simultaneous analysis of the causes of crime.

7 boys from the treatment group were dropped either because we possessed insufficient information on them or their behavior was so highly contradictory that they could not be placed in one of the three categories.

[3] One general issue deserves to be discussed at this point. In the next pages, we will report a number of significant relationships between the child's environment and his level of aggression. Can these relationships be attributed to a "halo effect"? Critics might reasonably argue that a specific association—say, between the father's level of aggression and his son's behavior—might be illusory and simply the result of the observers' preconceived biases. A counselor (or other observer), noting that a boy was highly aggressive, might unconsciously tend to report a number of "negative" observations about the child's environment, thereby falsifying the "real" situation.

For four reasons, however, we are convinced that the relationship between familial experiences and

2. The raters who, between 1956 and 1958, recategorized this information were also unaware that the data would be used for a study of aggression and were not acquainted with the hypotheses to be tested.

3. These ratings have previously been used in follow-up studies of those boys who became criminals, alcoholics, and psychotics in adulthood. The ratings successfully differentiated between these deviants and the nondeviant cases in the sample. Thus, in terms of certain forms of deviant behavior, the "predictive validity" of the categorizations has been established (cf. McCord, McCord, & Gudeman, 1960; McCord, McCord, & Zola, 1959).

aggression cannot reasonably be credited to a mere "halo effect":

A. As we have previously noted, the ratings of the home environment have proven predictive ability. In previous research, we have found that this set of ratings was strongly correlated with *adult* crime and *adult* alcoholism. Since the general predictive validity of these categorizations has been established, there would seem to be little reason to question their applicability to a study of aggression in childhood.

B. There is some empirical evidence that certain of the more obvious biases did *not* contaminate the data. One such bias is the possibility that middle class "Yankee" observers would tend to interpret behavior solely in terms of middle class standards. One might argue, for example, that a middle class person might, in watching lower class behavior, "read in" an aggressive intent, where none was intended—simply because the researcher was unacquainted with the customs of lower class culture. Similarly, an observer raised in an American milieu might tend to distort the purpose and significance of the actions of an Italian, Puerto Rican, or French-Canadian immigrant. We know that these possible biases did not significantly affect the data: there were no major differences in the ratings of aggression according to the social class of the child (of the middle class children, 16% were "aggressive," 56% were "assertive," and 28% were "nonaggressive"; of the lower class children, 12% were "aggressive," 57% were "assertive," and 31% were "nonaggressive"). Similarly, there were no significant differences in the ratings of aggression according to the child's ethnic group (of the sons of immigrants, 11% were rated as "aggressive," 55% as "assertive," and 34% as "nonaggressive"; among the sons of native Americans, 17% were regarded as "aggressive," 56% as "assertive," and 27% as "nonaggressive"). Further evidence against an operation of the "halo effect" is indicated by the fact that the staff did not rate the more intelligent boys in a positive fashion; equal proportions of intelligent and unintelligent boys were placed in each category of aggression.

C. Moreover, the original reports on both the boy and the home came from a variety of individuals (the typical case history included reports from an average of three counselors, four teachers, a psychiatrist, a physician, a psychologist, and various community officials); it would be remarkable if all of these people held the same unconscious biases.

D. And, as we have noted, the possibility that *conscious* biases pervaded the data was largely, perhaps entirely, eliminated. The original observers of the child and his environment did not know that their reports would be used for a study of aggression. (They were aware, of course, that the data would be used for a study of delinquency; but, as we have previously noted, known criminals have been eliminated from this study.) The later raters, who categorized the material in 1956, were also unaware that the information would be used to study aggression. Hence, the chance that either the original observers or the later researchers would consciously allow their own theories to contaminate the data was minimized.

RESULTS

We assumed, as the basic premises of this research, that early environmental experiences might affect a child's level of aggression in four conceptually distinct ways:

1. Emotional relationships between the boy and his parents would be fundamental in two ways: first, by affecting the child's level of frustration (and, thus, indirectly, his aggressive desires) and secondly, by molding his conception of the nature of human interaction.

2. The parental methods of disciplining and controlling the child would affect his willingness to inhibit whatever aggressive desires he felt.

3. The example of the parents (regardless of their direct relationship with the son) would affect the son by offering him an immediate, intimate model of human reactions to frustration.

4. The degree to which the parents reinforced each other's values would affect the "intensity" with which the son internalized parental demands (that is, we believed that parents who disagreed on a variety of issues, including the behavior which they expected of the child would be likely to produce a confused, relatively uninhibited child).

Let us examine the actual effects of each of these four general influences.

The Instigation of Aggression

Direct parental attacks on the boy—whether manifested in physically punitive discipline, the frequent use of threats, or constant unfavorable comments on the boy's worth—were strongly associated with a high level of aggressive behavior. Presumably, this kind of parental behavior is highly frus-

trating to the boy's dependency needs; parental rejection teaches him early in life that other human beings are threatening and aggressive. One indication of this relationship between parental punitiveness and aggression among the sons can be found in an examination of parental methods of discipline. Some of the parents employed "nonpunitive" means of disciplining their children; they reasoned with the son, withdrew certain privileges, or indicated their verbal disapproval. Other parents depended on directly "punitive" techniques: if the child disobeyed, they responded with angry abuse, slaps, or beatings. A significantly lower proportion of the nonaggressive boys than of the aggressive or assertive children were raised by "punitive" mothers. (Throughout the paper, the Scott π reliability coefficient was used to estimate interrater reliability. Reliability on judgments of maternal punitiveness was .772, on father's punitiveness, .581. A comparison of the nonaggressive boys to the others in terms of maternal punitiveness was significant at the .025 level; $\chi^2 = 5.09$; $df = 1$. A comparison of the effects of the fathers' techniques was not significant. Throughout the paper, the chi square test, two-tailed, was applied.)

Another indication that parental attacks on the child led to more aggressive behavior was that those parents who frequently threatened the child, who verbally frightened him with dire consequences if he disobeyed, were more likely to produce aggressive sons than were parents who rarely threatened the boy ($p < .05$; $\chi^2 = 6.11$; $df = 2$; Scott π reliability = .734).

As one might predict from the preceding evidence, those parents who generally disliked and rejected their sons were most likely to produce aggressive boys. In terms of their overall emotional relationship with the boy, the parents in the study were described as either "affectionate" or "rejecting." Affectionate parents were individuals who talked of their love for the son, who never voiced overtly rejecting statements about the son, and who demonstrated (at least, to some extent) their regard for the boy through kisses, hugs, gifts, proud statements about his accomplishments, etc. Rejecting parents, in contrast did not hesitate to express a fundamental dislike for their sons; they sometimes mentioned that they wished that the boy had never been born; they treated the boy either with open cruelty, neglect, or an erratic alternation

TABLE 1
PARENTAL METHODS OF DISCIPLINE AND THE CHILD'S AGGRESSION
(in percentages)

	Aggressive Boys	Assertive Boys	Nonaggressive Boys
Mother's method of discipline:	($N = 24$)	($N = 95$)	($N = 49$)
Punitive	54	48	31
Nonpunitive	46	52	69
	100	100	100
Father's method of discipline:	($N = 19$)	($N = 75$)	($N = 41$)
Punitive	58	49	41
Nonpunitive	42	51	59
	100	100	100

TABLE 2
PARENTAL USE OF THREATS AND THE CHILD'S AGGRESSION
(in percentages)

	Aggressive Boys ($N = 22$)	Assertive Boys ($N = 78$)	Nonaggressive Boys ($N = 41$)
Frequent use of threats by the parents	64	44	32
Infrequent use of threats by the parents	36	56	68
	100	100	100

TABLE 3
PARENTAL RELATION WITH THE BOY AND THE CHILD'S AGGRESSION
(in percentages)

Parental Attitudes Toward the Child	Aggressive Boys ($N = 19$)	Assertive Boys ($N = 78$)	Nonaggressive Boys ($N = 41$)
Affectionate mother and:			
affectionate father	5	60	68
rejecting father	37	19	17
Rejecting mother and:			
affectionate father	47	5	10
rejecting father	11	16	5
	100	100	100

between demonstrations of affection and blatant dislike.

As Table 3 demonstrates, the parents of the aggressive boys had an emotionally meager, cold relationship with their sons: The great majority of the aggressive boys, 95%, were raised in homes where at least one parent was emotionally rejecting; the majority of the assertive boys and of the nonaggressive children were raised by affectionate parents ($p < .001$; $\chi^2 = 22.28$; $df = 1$; Scott π reliability on mothers was .757, on fathers, .710).

The variables which we have discussed up to this point—parental rejection, punitiveness, and parental threats—have certain features in common. These attitudes and actions are all, presumably, direct attacks on the child's sense of security; they all tend to undermine the boy's conception of himself as a person of worth and significance; and they all carry the implication that the world is a dangerous, hostile environment. It would seem reasonable to suppose that these influences serve to arouse pervasive aggressive tendencies in the child; he is not only extremely frustrated by these attacks but he is implicitly taught that aggression is the "way of the world." If one views these variables in combination, an expected pattern appears: the aggressive children usually came from environments where several "drive-instigating" influences were present, while the assertive and nonaggressive children typically emerged from relatively "positive" homes.

TABLE 4

Drive-Instigating Variables and the Boy's Aggression[a]
(in percentages)

Number of "Negative" Factors	Aggressive Boys ($N = 25$)	Assertive Boys ($N = 97$)	Nonaggressive Boys ($N = 52$)
0	20	22	40
1	0	28	27
2	28	19	21
3	44	19	6
4	4	8	4
5	4	4	2
	100	100	100

[a] Each "negative" factor—maternal rejection, paternal rejection, maternal punitiveness, paternal punitiveness, and parental use of threats—was given a weight of one. If there was insufficient information, the factor was regarded as a "positive" element in the child's background.

Thus, 80% of the aggressive boys, 50% of the assertive boys, and only 33% of the nonaggressive boys had been exposed to two or more "negative" influences ($p < .001$; $\chi^2 = 15.12$; $df = 2$). It should be noted, in passing, that the delinquents (who were not included in Table 4) were usually exposed to at least three "negative" conditions.

In summary, this evidence leads to the conclusion that:

1. The aggressive boys and the assertive boys were more likely to have been disciplined in a punitive fashion by their mothers than were the nonaggressive boys ($p < .025$).

2. The aggressive boys were more likely to have been frequently threatened by their parents than were the assertive boys and the nonaggressive boys ($p < .05$).

3. The aggressive boys were more likely to have been raised by parents who rejected them than were the assertive or the nonaggressive boys ($p < .001$).

4. The three types of boys differed significantly in the total number of "aggression-instigating" elements in their backgrounds ($p < .001$).

The Direct Instillation of Controls

The preceding evidence suggests that a high level of aggression was originally instigated by parental rejection, punitiveness, and threats. Another general set of factors—the extent to which the parents exerted immediate controls over the boy's behavior—also played an important part. We assumed that such factors as the kind of demands imposed on the child, the degree to which his actions were supervised by the parents, and the extent to which the parents imposed their values in a consistent fashion could serve to modify the boy's expression of aggressive urges. Even a child who was plagued by intense aggressive urges, we assumed, might not express these desires overtly, if he was subjected to consistent controls by his parents. Several measures revealed the influence of this aspect of the boy's environment in the molding of aggressive behavior.

The parents differed strikingly in the kinds of demands which they imposed on their sons. In some families, the parents placed "high demands" on the child for conformity; the

son was expected to clean his room, do chores around the house, perform well in school, be polite, attend Sunday school—in other words, to conform to the dominant customs of American society. In other families, the parents did not require this high degree of "polite behavior." In some cases, they simply ignored the child; in other cases, they demanded a high level of performance in one area of behavior (e.g., wearing clean clothes) but gave the child complete freedom in other ways (e.g., behavior in the classroom).

The nonaggressive boys were more likely than either the assertive or the aggressive children to emerge from families that placed high demands on them for "polite," "responsible" behavior ($p < .005$; $\chi^2 = 9.09$; $df = 1$; Scott π reliability = .535).

As in the kind of demands imposed on the child, the families also differed in the degree to which they directly supervised the boy. Some of the parents (or other adults) kept a close eye on the boy's activities; they did not hesitate to advise the child or to intervene directly in behavior which they disapproved; they wished to know where he was and what he was doing most of the time. They were anxious to hear reports of his behavior away from the home. Other parents provided little, if any, supervision over the child; he was allowed to roam at will, to choose his own activities and friends. The nonaggressive children were significantly more likely to have been closely supervised ($p < .05$; $\chi^2 = 4.0$; $df = 1$; Scott π reliability = .710).

TABLE 5

PARENTAL DEMANDS ON THE CHILD AND THE BOY'S AGGRESSION
(in percentages)

Parental Demands	Aggressive Boys ($N = 26$)	Assertive Boys ($N = 96$)	Nonaggressive Boys ($N = 51$)
High	16	24	45
Low	84	76	55
	100	100	100

TABLE 6

SUPERVISION OF THE CHILD AND THE BOY'S AGGRESSION
(in percentages)

Supervision	Aggressive Boys ($N = 25$)	Assertive Boys ($N = 97$)	Nonaggressive Boys ($N = 52$)
Present	52	61	75
Absent	48	39	25
	100	100	100

TABLE 7

MOTHER'S CONTROL OF THE BOY AND BOY'S AGGRESSION
(in percentages)

Mother's Control	Aggressive Boys ($N = 25$)	Assertive Boys ($N = 96$)	Nonaggressive Boys ($N = 51$)
Overcontrolled	40	27	53
Normal control	16	40	31
Subnormal control	44	33	16
	100	100	100

Another measure, a rating of the mother's control of the child, was also related to the level of aggressive behavior. The boys' mothers were divided into three categories: those who "overcontrolled" their children (who demanded that the boy should be close to the mother at all times and submit completely to her direction); those who exerted "normal controls" over the child (women who were concerned with the boy's activities but gave him freedom in certain areas, such as the choice of playmates or the type of clothes which he would wear); and those who "subnormally" controlled the boy (such mothers frequently were neglecting or unconcerned with the child's behavior). The aggressive boys tended to be either "overcontrolled" or "subnormally controlled" by their mothers; the assertive boys were raised in approximately equal proportions by each type of mother; and the nonaggressive children were most often reared by mothers who "overcontrolled" or "normally controlled" them ($p < .005$; $\chi^2 = 15.46$; $df = 4$; Scott π reliability = .771).

One other element, the consistency of parental discipline, deserves attention. The majority of the parents in the study disciplined their children in an "erratic" fashion: for the same action, the child was sometimes severely punished while, at other times, he was allowed (or even encouraged) to continue in his behavior. In these inconsistent families, parental discipline was strongly affected by swings in mood or by the immediate situation. Other parents were relatively consistent; the child

could be sure that if he was caught committing a certain action, he would receive punishment. The nonaggressive children were raised by parents (particularly by mothers) who used consistent methods of discipline, while the converse tended to be true of the assertive and the aggressive boys (interrater reliability on ratings of parental consistency was .772, for the mothers, and .581 for the fathers. Comparison of nonaggressive boys to others in terms of consistency of maternal discipline was significant at the .01 level; $\chi^2 = 7.03$; $df = 1$; in terms of paternal consistency, the comparison was not significant).

It is clear that those variables which deal with direct parental control over the child—the kind of demands that they imposed on the son, the degree of supervision that they gave him, and the consistency of their discipline—have an important relation to the boy's aggression. These influences in the child's environment are only theoretically distinct from those factors which are related to the instigation of a high degree of aggression. In practice, for example, one would expect that parental rejection would usually be combined with a lack of direct parental supervision. It is useful, therefore, to view the impact of these conditions in combination. In Table 9, one can observe the interaction between one factor that seems to be primarily related to the instigation of aggression (parental punitiveness) and another factor that appears to be principally related to the instillation of controls (parental demands on the child).

TABLE 8
CONSISTENCY OF PARENTAL DISCIPLINE AND BOY'S AGGRESSION
(in percentages)

	Aggressive Boys ($N = 25$)	Assertive Boys ($N = 97$)	Nonaggressive Boys ($N = 52$)
Consistency of mother's discipline:			
Consistent	28	42	52
Erratic	72	58	48
	100	100	100
Consistency of father's discipline:			
Consistent	28	30	40
Erratic	72	70	60
	100	100	100

TABLE 9
PARENTAL PUNITIVENESS, PARENTAL DEMANDS ON THE CHILD, AND THE BOY'S AGGRESSION
(in percentages)

	Aggressive Boys ($N = 24$)	Assertive Boys ($N = 95$)	Nonaggressive Boys ($N = 45$)
The families imposed high demands on the child and:			
Both parents were punitive	4	3	8
One parent was punitive but the other parent was nonpunitive	8	14	18
Neither parent was punitive	4	6	19
The families imposed low demands on the child and:			
Both parents were punitive	13	21	8
One parent was punitive but the other parent was nonpunitive	54	34	14
Neither parent was punitive	17	22	33
	100	100	100

Both factors seem to be relevant to the expression of aggression. The aggressive boys tended to come from homes in which at least one parent was physically punitive and the demands on the child were low: 67% of the aggressive boys, 55% of the assertive children, and only 22% of the nonaggressive cases emerged from this type of family ($p < .001$; $\chi^2 = 17.74$; $df = 2$). Nonaggressive behavior appears to be associated with a lack of punitiveness on the parents' part *and* with the imposition of high demands. Of those children who came from homes in which both parents were nonpunitive and placed high requirements on their children's behavior, 56% were nonaggressive. At the opposite extreme (homes in which both parents were punitive and did not impose high demands), only 15% were nonaggressive ($p < .05$; $\chi^2 = 6.63$; $df = 2$). In general, if demands were high, a lack of aggressive behavior resulted; if demands were low, parental punitiveness tended to increase the level of aggressive behavior.

In summary, this material has indicated that:

1. The aggressive and the assertive children were more likely to have been raised by

parents who did not impose high demands on them than were the nonaggressive boys ($p < .005$).

2. The aggressive boys and the assertive boys were less likely to have been closely supervised by their parents than were the nonaggressive boys ($p < .05$).

3. More often than the assertive boys, the nonaggressive children were raised by mothers who "overcontrolled" them while the aggressive boys tended to be reared by mothers who either "overcontrolled" or "subnormally" controlled their sons ($p < .005$).

4. The nonaggressive boys (and to a lesser extent, the assertive boys) were more likely to have been disciplined by their mothers in a consistent fashion than were the aggressive children ($p < .01$).

5. The aggressive boys and the assertive boys were more likely than the nonaggressive boys to have been raised by parents who were both punitive and placed low demands on their sons ($p < .001$).

The Parental Model

It would seem reasonable to expect that highly aggressive parents would be likely to produce aggressive children. Presumably, a boy exposed to a father who consistently reacted with aggression to crises in his home, his work, or his neighborhood would tend to emulate the parental model. Surprisingly, the data do *not* confirm this "common-sense" prediction. Although there was a slight tendency for aggressive parents to produce aggressive children, this trend was not statistically significant at the 5% level.

Three measures—ratings of the father's aggression, his reaction to specific crises, and his role in the family—were used to estimate the effect of the parental model on the boy's behavior.

In terms of the same definition that was applied to their sons, the fathers in the study were divided into "aggressive," "assertive," and "nonaggressive" men. Fifteen percent of the aggressive boys were raised by aggressive fathers, but so were 18% of the assertive children, and 10% of the nonaggressive children. At the opposite extreme, 10% of the aggressive boys, as opposed to 6% of the assertive and 18% of the nonaggressive cases, were reared by nonaggressive fathers (Scott π reliability = .619).

Judgments were also made concerning the reactions of the fathers to specific crises: loss of a job, failure to gain a promotion, loss of "face" in the community, legal involvements, etc. Most of the fathers responded "realistically" by attempting to correct the situation in an effective fashion. Other men reacted in an "escapist" way: they went on a drinking or eating binge or withdrew into total passivity. A third set of fathers responded with overt "aggression" (verbal or physical attacks on the source of their frustration).

Although one might expect that the aggressive fathers would "teach" their sons to respond in a similar fashion, this did not occur: almost equal proportions of the aggressive, assertive, and nonaggressive boys were raised by fathers who responded aggressively to crises (17%, 16%, and 11%, respectively). "Realistic" fathers, on the other hand, tended to have nonaggressive children; 33% of the aggressive boys, 39% of the assertive, and 56% of the nonaggressive boys were raised by such men ($p < .05$; $\chi^2 = 3.84$; $df = 1$; Scott π reliability = .678).

The fathers differed, too, in terms of their role within the family. Some fathers were "dictators"; they made all the decisions in the family and ruled the home with an iron grip; others were passive, ineffectual men, who left the guidance of the family to the mother; and still others were "leaders"—they made the basic decisions for the family, but only after consultation. Although one might expect that the aggressive boys would come most often from "dictatorial" fathers, this did not prove to be the case. Nineteen percent of the aggressive boys, as compared to 8% and 7% of the assertive and nonaggressive boys were reared by dictatorial fathers; 38% of the aggressive boys, as opposed to 51% of the assertive and 55% of the nonaggressive were raised by fathers who played the role of the leader; the remainder came from passive fathers (Scott π reliability = .632).

One can draw a general conclusion from these data: paternal aggression did not significantly heighten the son's aggression; rather, it would appear that any form of paternal "deviance" whether expressed in

aggression, "escapism," or an eccentric role within the family—tended to be associated with a higher level of aggression in the son. The nonaggressive boys were more often reared by nonaggressive fathers who responded to crises in a realistic fashion and acted the role of "leader" within the family; one can assume, therefore, that these children were offered a parental model of responsibility and relatively stronger control over deviant impulses. The aggressive children often lacked such a paternal model of inner control; it would seem to be this deficiency in their background (rather than a specific example of aggression) that may have contributed to their unrestrained behavior.

Two other pieces of data support this interpretation. Many of the fathers in the study participated in various forms of deviant behavior (criminality, alcoholism, psychoses, desertion, sexual promiscuity, etc.). Only rarely was this behavior directly aggressive in nature. The deviant men, however, were relatively uninhibited, escapist, and irresponsible; they failed to offer their sons an example of moderation. The aggressive boys and the assertive cases were more often raised by such deviant fathers; 48% of the aggressive boys, 52% of the assertive cases, and 39% of the nonaggressive children were reared by deviant fathers ($p < .05$; $\chi^2 = 3.84$; $df = 1$; Scott π reliability = .827).

The strength of religion within the family may also have played a part. The nonaggressive boys were more often reared in homes which were "strongly" religious (that is, families in which the parents attended church at least once a week and gave other signs of devotion to their faith). Forty-eight percent of the nonaggressive children, as compared to 34% of the assertive and only 21% of the aggressive boys, were raised by strongly religious mothers ($p < .10$; $\chi^2 = 5.68$; $df = 3$; Scott π reliability for father = .856, for mother, .676). The specific religious affiliation of the family did not significantly affect the level of the boy's aggression. The strongly religious parents were people who, in general, valued inner control, conformity, "humility," and nonaggressiveness more than did those who were simply nominal members of a religious faith. Thus, one can assume that the strongly religious parents more often offered a model of inner control to their sons.

These data suggest that a parental model of conformity and inner control (whether expressed in realistic reactions to crises, the fulfillment of a mature role within the family, strong adherence to a religious faith, or a lack of obvious deviance) may be associated with a low level of aggressive behavior on the son's part. More strikingly, the material indicates that an aggressive parental model is, in itself, no more productive of aggression in the child, than are other "milder" forms of parental deviance, irresponsibility, or escapism.

Even these tentative generalizations should be further qualified. Within the sample, there is a general correlation between parental deviance and such other factors as parental rejection, punitiveness, and lack of control of the child. Thus, the slight association between parental deviance and the boy's aggression may well be simply a reflection of these other factors.

Parental Agreement on Values

One final set of conditions within the home, the amount of agreement between the parents concerning their basic values, has bearing on understanding the origins of aggression. In some homes, the parents were in constant conflict about each other's attitudes and actions or were fundamentally dissatisfied with their place in the family; in other words, such parents tended to undermine, rather than to reinforce the attitudes of one another. We assumed that a child raised in such an environment would tend to become confused concerning parental expectations and would be likely to "write off" parental demands. Thus, it would seem reasonable to assume that these children would be less governable by their parents, and, specifically, less inhibited in the expression of their aggressive urges.

The data support this interpretation. The parents' relation to one another, as well as their immediate relation to the child himself, had bearing on the boy's level of aggression. As a group the aggressive boys more often came from homes that were disordered by parental conflict, lack of respect of the parents for each other, disagreement within the family concerning methods of child rearing,

and antagonism between the child and his relatives. The aggressive boys were, therefore, often raised in a milieu in which one parent undermined (or directly attacked) the other; an environment which was not conducive to the establishment of a consistent set of values.

A variety of measures, all tapping the same phenomena, suggested that the aggressive children came from homes which were fundamentally disrupted by parental disagreements. The specific figures are presented in Table 10.

In summary, these conclusions can be drawn from the material:

1. The aggressive and the assertive children came more often from families characterized by "intense" conflict between the parents—constant bickering, recriminations, and, at times, overt physical attacks ($p < .005$; $\chi^2 = 9.16$; $df = 1$; Scott π reliability = .647).

2. The aggressive and the assertive children more often came from families in which one parent held the other in "low esteem." These were homes in which the parents expressed, with complete openness, their lack of respect for the other person ($p < .005$; $\chi^2 = 8.65$; $df = 1$; Scott π reliability = .833).

3. The aggressive boys more often came from homes in which both parents were dissatisfied with their role in life. In these families, the mothers complained of the burdens of being a housewife and mother and the fathers expressed their disgust with their job, their role in the home, and their relation to society ($p < .02$; $\chi^2 = 8.37$; $df = 2$; Scott π reliability = .679).

4. The aggressive boys were more likely to come from homes in which the parents disagreed concerning methods of child rearing ($p < .10$; $\chi^2 = 3.67$; $df = 1$; Scott π reliability = .620).

5. The aggressive boys were more likely to come from homes in which the parents were *not* demonstratively affectionate toward one another ($p < .05$; $\chi^2 = 4.64$; $df = 1$; Scott π reliability = .631).

Taken together, these various measures of the "milieu" of socialization reveal that the aggressive boys (and, to a lesser extent the assertive children) were reared in families that had been disordered by a high degree of parental conflict and antagonism. The nonaggressive boys, in contrast, more often emerged from environments in which the parents respected one another, agreed on basic issues (including the ways in which their sons should be raised), and were relatively satisfied with their lot in life.

TABLE 10
PARENTAL RELATIONSHIPS AND THE BOY'S AGGRESSION
(in percentages)

	Aggressive Boys	Assertive Boys	Nonaggressive Boys
Parental conflict:	($N = 21$)	($N = 83$)	($N = 43$)
Intense	38	36	12
Some	38	27	47
None	24	37	31
	100	100	100
Parental esteem:	($N = 20$)	($N = 81$)	($N = 38$)
High mutual esteem	50	52	79
One parent disrespected the other parent	25	20	13
Mutually low esteem	25	28	8
	100	100	100
Parental role satisfaction:	($N = 20$)	($N = 84$)	($N = 42$)
Both parents dissatisfied	45	26	12
One parent dissatisfied	5	25	19
Neither parent dissatisfied	50	49	69
	100	100	100
Parental conflict about the child:	($N = 21$)	($N = 76$)	($N = 41$)
Intense conflict	24	10	10
Some conflict	38	30	24
No conflict	38	60	66
	100	100	100
Parental affection for each other:	($N = 20$)	($N = 83$)	($N = 39$)
Affectionate	15	38	44
Sporadically affectionate	40	27	36
Indifferent	10	8	10
Antagonistic	35	27	10
	100	100	100

A natural effect of such parental conflict would be that the son tended to reject his family and the values that the parents represented. As one would expect, the aggressive boys more often chose a "reference group" (delinquent gang, neighborhood group, etc.)

outside of their home. Thirty-two percent of the aggressive children, as contrasted to only 14% of the assertive boys and 6% of the nonaggressive boys consciously oriented their behavior to an "outside" reference group rather than to their parents' wishes ($p < .02$; $\chi^2 = 6.58$; $df = 1$).

This tendency for the aggressive children to separate themselves from their families was also revealed in their attitudes towards their siblings. Only 42% of the aggressive children had a cooperative, friendly relationship with their siblings. In contrast, 70% of the assertive boys and 66% of the nonaggressive boys had an affectionate regard for their brothers and sisters ($p < .02$; $\chi^2 = 5.56$; $df = 1$).

CONCLUSIONS

In this paper, we have argued that three major elements in a child's environment affect his subsequent level of aggressive behavior:

1. The emotional relationship between the boy and his parents (the degree to which the family treats the child in a punitive, threatening, or rejecting fashion).

2. The direct controls exerted by the parents over the boy's behavior (the degree to which the parents guide the child in a consistent, rigorous fashion).

3. The pervasive milieu of the family (the extent to which the parents reinforced or undermined the values and significance of one another).

A fourth factor—the conformity or deviance of the parental model—may also be weakly associated with the child's level of aggression.

The data presented in this study—data gathered from the direct observation of 174 nondelinquent boys and their families over a period of more than 5 years—indicate that "aggressive," "assertive," and "nonaggressive" boys emerged from radically different environments.

These contrasts in background can perhaps best be viewed as summarized in Table 11.

These variables in the child's environment are not, of course, completely independent. We have treated particular conditions in the home, such as parental rejection of the child, as primarily related to the instigation of aggressive desires. Clearly, however, a factor such as parental rejection has other effects: it offers the child a model of aggression and it presumably weakens the child's desire to follow the demands of his parents. Certain

TABLE 11

BACKGROUNDS OF AGGRESSIVE, ASSERTIVE, AND NONAGGRESSIVE BOYS

Environmental Conditions	Aggressive Boys	Assertive Boys	Nonaggressive Boys
Parents' emotional relationship with the boy	Rejecting	Affectionate	Affectionate
	Punitive	Relatively punitive[a]	Nonpunitive
	Frequent use of threats	Infrequent use of threats	Little use of threats
Instillation of direct controls	Overcontrolled or subnormally controlled by mothers	Normally or subnormally controlled by mothers	Normally or overcontrolled by mothers
	Low demands on the child	Low demands	High demands
	Lack of supervision	Relatively little supervision of the child[a]	Firm supervision of the child
	Inconsistency in discipline	Relatively inconsistent in discipline[a]	Consistent in discipline
Parental model	Socially deviant	Relatively deviant[a]	Socially conformist
Parental relationship to each other	A high degree of general conflict	High degree of general conflict	Low degree of general conflict
	Lack of mutual esteem	Relatively low mutual esteem[a]	High mutual esteem
	Dissatisfaction with role in life	Relatively dissatisfied with role[a]	Satisfied with role
	Unaffectionate	Affectionate	Affectionate

[a] In comparison to the background of nonaggressive boys.

other variables, however, do not seem to have this many-faceted influence. The degree to which the child has been subjected to adult supervision, for example, would seem to have direct relevance to controlling the boy's expression of aggression, but little or no bearing on the original establishment of aggressive desires.

To construct a comprehensive picture of the origins of aggression, one must view these various conditions in the child's environment in terms of their total impact. In an attempt to develop a general portrait of the origins of aggression, 10 variables in the background of every child in the study were categorized.

First, the children were divided according to the number of "aggression-creating" factors in their environment. For this purpose, five variables were used: paternal attitude toward the boy, maternal attitude toward the child, paternal punitiveness, maternal punitiveness, and parental use of threats. For each factor that was present in the child's home, a "negative" score of one was given. Thus, the child who was rejected by both parents, who was treated punitively by both parents, and was subjected to frequent threats received a score of "−5." At the opposite extreme, the child who had not been subjected to any of these influences received a score of "0." If we did not possess sufficient information on a particular variable, this factor in the child's background was, *de facto*, considered as "positive."

Secondly, five variables relating to the general milieu of the home, the parental model, and the controls exercised by the parent over the child were considered—in other words, those factors which were presumably related to modifying the expression of aggressive desires. These conditions included parental supervision of the child, maternal consistency of discipline, parental demands on the boy, the strength of the maternal religion, and the level of general conflict between the parents. Each factor was given a "positive" weight of one; thus, the boy who came from an environment characterized by maternal consistency, firm supervision, a strong religious faith, high demands, and parental agreement was awarded a score of "+5." If we did not possess sufficient information on a particular condition, the factor was considered as "negative."

Since we knew that each of these 10 variables was independently related to aggression, our supposition was that their combined influence would account for much of the variation in aggressive behavior. We assumed, in particular, that the aggressive boys would be most likely to emerge from an environment that simultaneously instilled a high level of aggressive urges and failed to provide the conditions for controlling aggression. The nonaggressive boys, in contrast, would be expected to come from homes characterized by a high level of controls and a low degree of aggression-creating influences. As Table 12 indicates, these expectations were generally fulfilled.

In a general sense, therefore, we can argue that the *aggressive boys* were most likely to have been raised by parents who (*a*) treated the boy in a rejecting, punitive fashion; (*b*) failed to impose direct controls on his behavior; (*c*) offered him an example of deviance; and (*d*) were often involved in intense conflict.

The *nonaggressive boys* came from a strikingly contrasting environment—a home in which they were (*a*) treated in an affectionate,

TABLE 12

THE COMBINED IMPACT OF ENVIRONMENTAL CONDITIONS ON CHILDHOOD AGGRESSION
(in percentages)

	Aggressive Boys ($N = 25$)	Assertive Boys ($N = 97$)	Nonaggressive Boys ($N = 52$)
An environment in which the parents imposed "high" controls over the son's behavior (+ 3, + 4, and + 5) and the home was:			
Highly conducive to aggression (−5, −4)	4	1	2
Moderately conducive to aggression (−3, 2)	4	11	13
Not conducive to aggression (−1, 0)	8	24	39
An environment in which the parents imposed few controls over the son's behavior (+ 2, + 1, and 0) and the home was:			
Highly conducive to aggression (−5, −4)	4	11	4
Moderately conducive to aggression (−3, −2)	68	26	15
Not conducive to aggression (−1, 0)	12	27	27
	100	100	100

nonpunitive manner; (*b*) guided by a consistent set of controls; (*c*) exposed to examples of social conformity; and (*d*) reared by affectionate, satisfied parents.

The *assertive boys* generally resembled the nonaggressive children in that they were reared by relatively affectionate, nonthreatening parents; thus, they were not victimized by the rejection characteristic of the families of aggressive boys (51% came from families which were nonpunitive and affectionate as opposed to only 20% of the aggressive boys). They resembled the aggressive boys, however, in that their parents (*a*) often failed to impose consistent controls; (*b*) were deviant models; and (*c*) often were in open conflict. (64% of the assertive boys came from an environment where these conditions were present.)

It would appear, therefore, that the aggressive boys had a high level of aggressive urges and were uncontrolled; the nonaggressive boys had a low degree of aggressive desires and, in addition, were well-controlled; the assertive boys had a few basically aggressive urges but were relatively uncontrolled individuals.

It is encouraging to note that the results of this study, despite methodological differences, are in fundamental agreement with the conclusions which have emerged from prior research. For example, Bandura and Walters (1959) have studied the childhood antecedents of adolescent aggression. Their sample, drawn from a group of middle class, antisocial, delinquents differed strikingly from the boys studied in this research (in age, in social class, in geographic area, and in illegal behavior). Yet the results are basically similar; the antisocial, aggressive boys came from homes in which they were rejected and treated in an inconsistent fashion.

Again, the results achieved by Sears, Maccoby, and Levin (1957) in their recent investigations of childhood aggression bear a basic resemblance to the findings of this research. Among other relationships, Sears and his colleagues found that aggression in childhood was associated with the use of physical punishment, low esteem of the mother for the father, a high degree of permissiveness for the expression of aggression, disagreement between the parents, and dissatisfaction on the mother's part with her role in life. And, indeed, these same relationships appeared in this present study—even though the sample used in this research came from a generally lower social class than did the Sears' sample and the methods of observation in the two studies were quite different.

These various relationships between the nature of a child's family and his level of aggressive behavior could be interpreted in at least four different ways:

1. Some analysts might contend that aggression on the child's part actually provokes certain "typical" responses from the parents. By this interpretation, the child's innate aggressive behavior evokes rejecting attitudes, punitiveness, and inconsistency from the parents; the behavior itself is regarded as creating new conditions within the family. While it seems highly probable that the child's behavior reciprocally affects parental attitudes towards him, it would be stretching this interpretation unduly to argue that the child's behavior also causes such diverse phenomena as conflict between the parents, lack of supervision, parental deviance, and low esteem on the parents' part for each other.

2. Some might argue that the parents' behavior and the child's aggression are produced by common genetic factors and that the apparent environmental relationships are illusory. Within the confines of this study, it is impossible to evaluate this opinion directly. In the absence of definitive genetic evidence, however, it seems most reasonable to view this interpretation as a rather remote possibility.

3. Some social scientists, particularly sociologists, might well contend that both the parents' and the boys' behavior result from similar vectors in the social environment. These analysts would place special emphasis upon the similar experiences undergone by the parents' and the boy in their relations with the "outside" world: the school, the occupational environment, the neighborhood, or the ethnic culture. Since there was no independent relation between the child's level of aggression and the family ethnic group or social class, this explanation does not appear to do justice to the findings.

4. Although the previous interpretations

cannot be dismissed, it seems most reasonable to assume that the familial environment is an independent variable affecting the child's behavior. The evidence presented in this study (and the facts accumulated by previous work) suggest that aggression in childhood is a form of behavior developed in response to specific environmental conditions; conditions created by man and thus, potentially changeable by man. Clearly, aggression is a universal capacity of human nature, a capacity which is first expressed in the unfocused rage of infants. But the development of this trait—whether it is transformed into a pervasively destructive syndrome of behavior or whether it goes fallow—seems to lie well within the realm of human culture, as this culture is mediated through early familial experiences.

REFERENCES

BANDURA, A., & WALTERS, R. *Adolescent aggression.* New York: Ronald, 1959.

GLUECK, S., & GLUECK, E. *Unraveling juvenile delinquency.* New York: Commonwealth Fund, 1950.

MCCORD, W., MCCORD, J., & GUDEMAN, J. *Origins of alcoholism.* Stanford: Stanford Univer. Press, 1960.

MCCORD, W., MCCORD, J., & ZOLA, I. *Origins of crime.* New York: Columbia Univer. Press, 1959.

POWERS, E., & WITMER, H. *An experiment in the prevention of delinquency.* New York: Columbia Univer. Press, 1950.

SEARS, R., MACCOBY, E., & LEVIN, H. *Patterns of child rearing.* Evanston, Ill.: Row, Peterson, 1957.

WEISS, W., & FINE, B. The effect of induced aggressiveness on opinion change. *J. abnorm. soc. Psychol.*, 1956, **52**, 109–114.

(Received December 15, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 94-98

SELF-CONCEPT DIFFERENCES IN RELATION TO IDENTIFICATION, RELIGION, AND SOCIAL CLASS[1]

JAMES BIERI AND ROBIN LOBECK[2]
New York School of Social Work, Columbia University

If the self-concept is a product primarily of the social learning of the individual, as most personality theorists assert, then it should be possible to demonstrate systematic differences in the self-concepts of persons whose social experiences have differed in fundamental ways. We have focused in this study on three variables that should reflect differential social learning experiences, in an exploratory attempt to determine how these variables relate to two aspects of the self-concept, *dominance* and *love*. Dominance includes reported self-behaviors involving assertion, control over others, and self-confidence, while love includes characteristics such as friendly, warm, and cooperative.

Most generally, males might be expected to develop differing self-concepts depending upon whether their identification is primarily with the father or the mother. If we consider dominance and love behaviors as cultural modal points in the role behavior of fathers and mothers, respectively, we would expect father identifiers to emphasize self-conceptions of dominance and mother identifiers to emphasize self-conceptions of love. That parental behavior may cluster about these modal points is suggested by our previous work (Bieri & Lobeck, 1959). Older adolescents tended to perceive their parents in terms of an active, controlling, and teaching role (dominance) or in terms of a role emphasizing emotional closeness and warmth (love). Similarly, the love and dominance dimensions of this study resemble, respectively, the love-hate and control-autonomy dimensions described by Schaefer (1959) in his factor analyses of results from parent behavior studies.

A second major socializing influence on how a person's self-concept develops is his religious affiliation. Religion may exert an influence on one's behavior by defining the available social and economic opportunities, by structuring certain values and attitudes, and by structuring social relationships within the family. We have selected two religious faiths, Catholicism and Judaism, about which, from the existing literature, one might assume certain differences in socialization that should affect the self-concept. Perhaps most crucial in terms of the dominance aspect of the self-concept are the value differences often attributed to these two groups (Sklare, 1958; Strodtbeck, 1958). A strong value in the Jewish group has been considered to be the emphasis upon the rational mastery of one's environment with concomitant emphasis upon achievement and intellectual attainment. In contrast, values associated with the Catholicism of immigrant groups from Ireland and Italy have included an emphasis on fatalism and the role of destiny in controlling one's life, as well as dependence upon external authority. Such considerations are akin to Florence Kluckhohn's categories of value orientations in which the Jewish group would be considered to stress man's mastery over nature while the Catholic group stresses man's subjugation to nature (Kluckhohn, 1953).

The third variable assumed to affect the self-concept is social class status. It seems reasonable to expect that the degree of dominance in one's self-concept varies directly with one's social class background. Persons from lower class backgrounds have been exposed to more learning conditions in which the relative inferiority of their physical and social situation may be translated into self-feelings of inferiority and submissiveness. Conversely, the upper status individual has been associated with conditions of occupational and social prestige and power, promoting a greater feeling of self-confidence and dominance.

METHOD

Subjects. Members of an enlisted men's Army reserve unit served as subjects for the research. This

[1] This study was supported by Research Grant M-1976 C1 from the National Institutes of Mental Health, United States Public Health Service.

[2] We wish to acknowledge the assistance of Audrey Laibson and Harold Plotnick in conducting this study.

group of 89 males, residents of the Bronx, ranged in age from 19 to 40 years, with a mean age of 23.5 years. The majority of the subjects identified themselves as either Jewish or Catholic. Approximately one-third of each religious group were married. The ethnic origin of the Catholic group was mixed, being predominantly Italian or Irish, while the Jewish group had predominantly Eastern European origin. Approximately 50% of the subjects in each religious group had fathers who were foreign born.

Self-concept. The two aspects of the self-concept, dominance and love, were measured by the Interpersonal Checklist (Leary, 1957). This method consists of 128 adjectives or phrases grouped into eight behavioral categories or octants. These octants are: managerial-autocratic, competitive-narcissistic, aggressive-sadistic, rebellious-distrustful, self-effacing–masochistic, docile-dependent, cooperative-overconventional, and responsible-hypernormal. The subject was instructed to check those adjectives or phrases which best described himself. Dominance and love scores were obtained by means of formulae developed by LaForge, Leary, Naboisek, and Coffey (1954). The dominance score is most positively weighted on the managerial-autocratic and competitive-narcissistic octants, and negatively weighted on the self-effacing–masochistic octant. The love score is most positively weighted on the cooperative-overconventional and docile-dependent octants, and negatively weighted on the aggressive-sadistic octant. Dominance and love scores were converted to standard scores with a mean of 50 and a standard deviation of 10. These two scores are not correlated ($r = -.03$) for the subjects in this study.

Identification. The measure of identification used here is intended to reflect the subject's perception of greater similarity to one parent than to the other. A more detailed analysis of this measure of one aspect of identification and its personality correlates may be found in previous publications (Bieri, 1960; Bieri & Lobeck, 1959; Bieri, Lobeck, & Galinsky, 1959). Each subject rated himself, his mother, and his father on 14 10-point scales between each of the following bipolar adjectives: carefree-serious, dependable-undependable, passive-active, happy-sad, delicate-rugged, loud-quiet, friendly-unfriendly, meek-bossy, unpleasant-pleasant, strong-weak, cold-warm, confident-unsure, excitable-calm, and smart-dumb. The perceived similarity to each parent is measured by squaring the distance between one's self-rating and the rating given the parent on each 10-point scale. The sum of these 14 distance scores (D^2) represents an index of the perceived similarity to each parent, and the subject is assumed to identify with (i.e., perceive himself as more like) that parent with whom he has the lower D^2 score. The independence of this identification measure and the checklist measure of the self-concept will be considered below.

Social class. The social class background of each subject was determined on the basis of Hollingshead's two-factor index (Hollingshead, 1957). This index classifies the subject into one of five categories on the basis of the father's occupation and education. The numerical scheme used places greater weight upon the father's occupation than upon his education. The 84 subjects used in the analyses involving social class were distributed into five classes as follows: Class I (highest), 3; Class II, 7; Class III, 18; Class IV, 42; and Class V, 14. Thus, the modal class of the sample is Class IV, the lower middle class. While the class distribution of the Jewish subjects indicated somewhat higher class backgrounds than those of the Catholic group, this difference was not significant. We have grouped these five classes into two overall class groupings, which we call the upper class group (I, II, and III) and the lower class group (IV and V).

RESULTS AND DISCUSSION

Table 1 presents the mean scores on dominance and love as a function of the three socialization variables. Considering identification first, no significant differences within or between identification groups are found. However, the differences are in the direction anticipated in our earlier discussion, with father identifiers having higher dominance than love scores, and mother identifiers having higher love than dominance scores. The most striking difference is the higher love score of mother identifiers compared to father identifiers ($p = .11$).

When the results in Table 1 are analyzed in relation to religious affiliation, it is found that while Jewish and Catholic subjects have almost identical mean dominance scores, Catholic subjects have a significantly higher mean love score than do Jewish subjects. Within religious groups, there is not a significant difference between mean dominance and love scores for Jewish subjects, but Catho-

TABLE 1

MEAN DIFFERENCES IN DOMINANCE AND LOVE SELF-CONCEPT SCORES IN RELATION TO PARENTAL IDENTIFICATION, RELIGIOUS AFFILIATION, AND SOCIAL CLASS STATUS

Subgroup	*N*	Dominance	Love	*p*
F	39	49.62	48.59	*ns*
		p *ns*	*ns*	
M	47	49.08	52.49	*ns*
J	48	49.79	48.08	*ns*
		p *ns*	.02	
RC	41	49.83	53.90	.01
I–III	28	54.14	47.96	.05
		p .01	*ns*	
IV–V	56	47.23	51.34	*ns*

Note.—In this and subsequent tables, F = father identifiers, M = mother identifiers, J = Jewish subjects, RC = Catholic subjects, I-III = upper class subjects, and IV-V = lower class subjects. All probability values are based upon *t* tests using a two-tailed test. Incomplete data on some subjects causes the *N*s to vary.

lic subjects have a significantly higher mean love score than mean dominance score. Both significant findings in relation to religious affiliation involve the elevated love scores in the Catholic group.

Finally, in relation to social class, we observe in Table 1 that upper class subjects do have a significantly higher mean dominance score than lower class subjects. While lower class subjects have a higher mean love score than upper class subjects, this difference is not significant. Within class groups, upper class subjects have a higher mean score on dominance than on love, and lower class subjects have a higher mean score on love than on dominance. Only the former difference is significant, however.

Further analyses of these self-concept differences are possible holding the effects of two of the variables constant when studying the third. However, a problem arises in this regard because of the lack of independence among many of the comparisons that could be made. To avoid having possible a posteriori judgments dictate what independent comparisons are made in an exploratory study such as this, we present and discuss these subgroup findings as suggestions for further study in which a priori considerations would indicate the specific relationships to be tested.

Since our self-concept measure and our identification measure both contain self-ratings, it is necessary to analyze the possible confounding of these two measures. As would be expected, there is some association between the self-ratings on the checklist and the self-ratings on the semantic differential. If we isolate a "warmth" cluster of three scales on the semantic differential (warm-cold, friendly-unfriendly, and pleasant-unpleasant), self-scores on this cluster correlate significantly with the love scores ($r = .48$, $p < .01$). Similarly, self-scores on a "strength" cluster of the semantic differential scales (strong-weak, bossy-meek, confident-unsure, and rugged-delicate) correlate significantly with the dominance scores ($r = .47$, $p < .01$). Moreover, item analysis shows these "warmth" and "strength" scales to be the best predictors of the overall direction of identification. Thus, for example, on the warm-cold scale mother identifiers score themselves closer to the mother than to the father while father identifiers score themselves closer to the father than to the mother (chi square = 17.66, $p < .01$). That is, identification as we measure it appears to reflect to a great extent perceived similarities in terms of warmth and strength. Now, we also find that subjects in both identification groups tend to rate mother higher than father on the "warmth" scales, and father higher than mother on the "strength" scales. Does it follow, then, that subjects who score closer to the mother than to the father (mother identifiers) rate themselves higher on "warmth" scales and lower on "strength" scales than subjects who score closer to the father than to the mother (father identifiers)? If so, it is obviously redundant to say that the subject perceives himself to be high on love *and* that he identifies with the parent who is considered to be more loving (the mother). However, such a confounding of identification and self-concept reouires that it is the self-ratings on the semantic differential and not the ratings given the two parents which differentiate mother from father identifiers. When we examine the relative mean ratings given the mother, father, and self in the two groups, we find that on the scales of the "warmth" cluster it is the difference in the rating given the *father* which distinguishes mother from father identifiers. On the "strength" scales, the two identification groups are distinguished by their positioning of the *mother*. In neither cluster does a difference in self-rating distinguish between mother and father identifiers. Thus, even though the self-concept is involved in obtaining the measure of identification, identification per se does not predict the subject's self-concept.

As noted above, the significant findings relative to religious affiliation (Table 1) center around the elevated mean love score in the Catholic group. Reference to Table 2 indicates that in every subgroup comparison between subjects in the two religious groups, Catholic subjects have higher love scores than do Jewish subjects. However, within the Catholic group, it appears that the mother identifiers are contributing most to these differences in love self-conceptions. Thus, the more numerous Catholic mother identifiers have a mean love score of 55.56, while the less numerous Catholic father identifiers have a mean love score of 50.60. Further, the mean love scores

TABLE 2

MEAN SELF-CONCEPT SCORES CLASSIFIED BY RELIGION, IDENTIFICATION, AND SOCIAL CLASS BACKGROUND

Subgroup	Dominance	Love	Social Class I–III Dominance	Social Class I–III Love	Social Class IV–V Dominance	Social Class IV–V Love
All subjects	49.80 (N = 89)	50.76	54.14 (N = 28)	47.96	47.23 (N = 56)	51.34
F	49.62 (N = 39)	48.59	57.23 (N = 13)	45.92	45.38 (N = 25)	49.48
M	49.08 (N = 47)	52.49	50.93 (N = 14)	49.43	47.45 (N = 29)	52.97
J	49.79 (N = 48)	48.08	55.16 (N = 19)	46.00	46.00 (N = 28)	49.04
JF	51.04 (N = 24)	47.33	57.70 (N = 10)	45.30	45.69 (N = 13)	47.85
JM	47.50 (N = 22)	49.00	51.50 (N = 8)	45.87	45.21 (N = 14)	50.79
RC	49.83 (N = 41)	53.90	52.00 (N = 9)	52.11	48.46 (N = 28)	53.64
RCF	47.33 (N = 15)	50.60	55.67 (N = 3)	48.00	45.25 (N = 12)	51.25
RCM	50.48 (N = 25)	55.56	50.16 (N = 6)	54.16	49.53 (N = 15)	55.00

TABLE 3

CHECKLIST ITEMS SELECTED WITH DIFFERENTIAL FREQUENCY BETWEEN SUBGROUPS ON EACH SOCIALIZATION VARIABLE

		p
Identification		
wants to be led	F > M	.02
Religion		
can be indifferent to others	J > RC	.01
critical of others	J > RC	.02
able to doubt others	J > RC	.05
able to criticize self	J > RC	.05
easily embarrassed	RC > J	.06
very respectful to authority	RC > J	.01
dependent	RC > J	.05
will believe anyone	RC > J	.05
friendly	RC > J	.10
fond of everyone	RC > J	.01
likes everybody	RC > J	.06
loves everyone	RC > J	.05
forgives anything	RC > J	.02
enjoys taking care of others	RC > J	.06
Social Class		
able to give orders	I–III > IV–V	.06
critical of others	I–III > IV–V	.09
businesslike	I–III > IV–V	.01
outspoken	I–III > IV–V	.05
passive and unaggressive	IV–V > I–III	.10
very respectful to authority	IV–V > I–III	.10
lets others make decisions	IV–V > I–III	.10

Note.—Probability values based upon chi square corrected for continuity, two-tailed test.

in the Catholic mother identification group are quite similar for subjects from lower class and upper class backgrounds. In this regard, it is interesting to observe (see Table 2) that in the three-way breakdown of religion, identification, and social class among upper class subjects only the Catholic mother identifiers have higher love than dominance scores.

Further analyses of the significant findings in relation to social class indicate that in both identification groups and in both religious groups, upper class subjects have higher dominance scores than lower class subjects (see Table 2). However, these social class differences in dominance are greater within either religious group for father identifiers than for mother identifiers. This finding suggests that identification may affect self-conceptions in interaction with other socialization variables.

An analysis of the specific checklist items that were selected more frequently by one group than the other within each of the socialization categories is presented in Table 3.

Catholic subjects checked more frequently than Jewish subjects items from the cooperative-overconventional octant of the checklist (fond of everyone, likes everybody, loves everyone, and friendly). This octant is weighted most heavily in the derivation of the love score. Three of these four items were selected significantly more often by Catholic mother identifiers than by Jewish mother identifiers. While 9 of the 10 items checked more frequently by Catholic subjects come from the cooperative-overconventional, docile-dependent, and responsible-hypernormal octants, none of the items checked more frequently by Jewish subjects comes from these octants. Rather, the four items checked more frequently by Jewish subjects represent the competitive-narcissistic, aggressive-sadistic, rebellious-distrustful, and self-effacing–masochistic octants. These item differences between the religious groups suggest more critical, skeptical, and autonomous self-conceptions in Jewish subjects and more conforming, nurturant, and acquiescent self-conceptions in Catholic subjects, especially Catholic mother identifiers.

The items checked more frequently by upper class subjects (see Table 3) come from the managerial-autocratic, competitive-narcissistic, and rebellious-distrustful octants, while lower class subjects more frequently check items from the self-effacing–masochistic and docile-dependent octants. These self-concept

differences between the two class groupings are consistent with the findings of other studies. Thus, Klausner (1953) reports more feelings of inferiority and insecurity in adolescents from lower class backgrounds. The greater emphasis on submissive and passive self-conceptions in our lower class subjects is consistent with the findings of Hollingshead and Redlich (1958) and others that schizophrenic symptoms, which may be extreme forms of these submissive trends, are perhaps more frequent in lower than upper class persons. Similarly, Myers and Roberts (1959) report that family relationships and life conditions appeared to be more conducive to the development of isolation and withdrawal in the Class V neurotics and schizophrenics of their sample. Lane and Singer (1959) also report that in contrast to the middle class normal subjects, lower class normals and schizophrenics evidence a greater amount of submission and denial of aggression.

SUMMARY

This study focused on the relationships between the three socialization variables of parental identification, religious affiliation, and social class to two aspects of the self-concept, dominance and love. The subjects in the study were 89 male Army reserve enlisted personnel who were about equally divided by Catholic or Jewish religious affiliation. Parental identification was measured in terms of the greater perceived similarity to one parent in contrast to the other on the semantic differential. Upper class background was measured by grouping Social Classes I through III on Hollingshead's two-factor index, while lower class background consisted of Classes IV and V on the same index. The primary findings were:

1. No significant differences in self-concept scores as a function of parental identification were observed.

2. Catholic subjects have significantly higher love scores than do Jewish subjects, and the former have significantly higher love than dominance scores.

3. Upper class subjects have significantly higher dominance scores than lower class subjects, and upper class subjects are significantly higher on dominance than on love.

Further analyses of the data suggest that future research might study systematically the suggestive findings that (*a*) love self-conceptions are most pronounced in Catholic mother identifiers and that (*b*) the greatest differences in dominance self-conceptions are found between groups of father identifiers who differ in terms of social class status.

REFERENCES

BIERI, J. Parental identification, acceptance of authority, and within-sex differences in cognitive behavior. *J. abnorm. soc. Psychol.*, 1960, **60**, 76-79.

BIERI, J., & LOBECK, ROBIN. Acceptance of authority and parental identification. *J. Pers.*, 1959, **27**, 74–86.

BIERI, J., LOBECK, ROBIN, & GALINSKY, M. D. A comparison of direct, indirect, and fantasy measures of identification. *J. abnorm. soc. Psychol.*, 1959, **58**, 253–258.

HOLLINGSHEAD, A. B. Two factor index of social position. Author, 1957.

HOLLINGSHEAD, A. B., & REDLICH, R. C. *Social class and mental illness*. New York: Wiley, 1958.

KLAUSNER, S. J. Social class and self-concept. *J. soc. Psychol.*, 1953, **38**, 201–205.

KLUCKHOHN, FLORENCE, R. Dominant and variant value orientations. In C. Kluckhohn, H. A. Murray, & D. M. Schneider (Eds.), *Personality in nature, society, and culture*. (2nd ed.) New York: Knopf, 1953. Pp. 342–357.

LA FORGE, R., LEARY, T., NABOISEK, H., & COFFEY, H. The interpersonal dimension of personality: II. An objective study of repression. *J. Pers.*, 1954, **23**, 129–153.

LANE, R. C., & SINGER, J. L. Familial attitudes in paranoid schizophrenics and normals in two socioeconomic classes. *J. abnorm. soc. Psychol.*, 1959, **59**, 328–339.

LEARY, T. F. *Interpersonal diagnosis of personality*. New York: Ronald, 1957.

MYERS, J. S., & ROBERTS, B. H. *Family and class dynamics in mental illness*. New York: Wiley, 1959.

SCHAEFER, E. S. A circumplex model for maternal behavior. *J. abnorm. soc. Psychol.*, 1959, **59**, 226–235.

SKLARE, M. (Ed.) *The Jews*. Glencoe, Ill.: Free Press, 1958.

STRODTBECK, F. L. Family interaction, values, and achievement. In D. C. McClelland, A. L. Baldwin, U. Bronfenbrenner, & F. L. Strodtbeck (Eds.), *Talent and society*. Princeton: Van Nostrand, 1958. Pp. 135–194.

(Received December 17, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 99-105

EMOTIONAL FACTORS IN THE FORGETTING OF WORD ASSOCIATIONS[1]

GEORGE LEVINGER
Western Reserve University

AND

JAMES CLARK[2]
Michigan State University

SINCE the publication of Jung's pioneering studies on word association (1906), there has been a steadily increasing number of studies directed toward the possible relationship between emotional and cognitive variables. In the 1920s and 1930s, a major research effort was to demonstrate that emotion is one of the determinants of memory. Apparently, this effort was part of a more general attempt to examine the implications of statements by Freud, Jung, and other psychoanalytic writers. Rapaport (1942), Meltzer (1930), Zeller (1950), among others, have reviewed many of the studies which served either to support or to invalidate the above thesis, and these studies will not be reviewed here. It might be noted, however, that in recent years interest in the emotional determinants of memory has diminished, and attention has shifted to affective factors related to perception. One sign of this shift was Rapaport's decision in 1950 to republish "unaltered" the 1942 edition of his book *Emotions and Memory*.

One might believe that this multitude of investigations would have led to some reasonably secure conclusions about the influence of emotional factors in the memory process. Yet this is not the case. For example, Bugelski's (1956) text on learning rejected the idea that there is any evidence linking emotional factors to memory. Underwood, in a recent statement, refused to attach importance to the role of affect in forgetting until further research establishes a more "reliable body of facts" (1957, p. 57). Even Rapaport (1942), after his rather sympathetic survey of the relevant literature, concluded that considerably more work is needed before one can understand the nature of such a relationship.

The present study grew out of various failures—by the first author and by other investigators (e.g., Grummon & Butler, 1953; Merrill, 1952)—to confirm the validity of a word association technique used by Keet (1948) for evoking the forgetting of "traumatic" words. Keet had employed the Jung association method for trying to locate that word in the stimulus list which resulted in the most "traumatic" response on the part of the subject (*S*). Having located such a word, Keet found that *S*s blocked in remembering it during a subsequent learning task, enabling Keet to introduce either of two experimental quasi-counseling methods for helping *S* to overcome his forgetting of the word.

After failing with 10 *S*s to replicate Keet's method for producing the supposed trauma, the first author tried a variation of his technique. Instead of asking *S*s merely to associate any word upon the second presentation of the stimulus list—which is the traditional procedure (e.g., Hull & Lugoff, 1921; Jung, 1906; Keet, 1948)—*S*s were asked to *remember* their first free association (the procedure used by Rapaport, Gill, & Schafer, 1946). Responding to this, *S*s recalled the large majority of their previous associations, but forgot many of those which had an apparently emotional significance. For example, one *S* responded "bad" to the stimulus word "love," and later failed to remember this response. To develop this procedure for the purpose of setting up a laboratory analogue of a counseling session though, it was necessary to demonstrate that the forgetting observed is indeed related to emotional factors.

The study reported here was conducted, therefore, to examine the general hypothesis that the forgetting of word associations is related positively to their emotional significance. In other words, it was expected that the more emotional the reaction to a stimulus, the

[1] The research reported here was done while the authors were at the University of Michigan. The study was supported, in part, by a grant (M-450) from the National Institute of Mental Health, United States Public Health Service, administered by E. S. Bordin. Portions of this paper were presented at the 1958 APA meetings in Washington, D. C.

[2] The authors are indebted to R. S. Davidon, R. R. Bush, and H. B. Ranken for commenting on certain statistical issues.

less likely that the specific response would be recalled.

This hypothesis seems relevant to other research on affective factors in recall. Yet the only other study known to the writers that dealt with directly pertinent data is one by Rapaport, Gill, and Schafer (1946). The writers reported that *Ss* forgot their associations to traumatic stimulus words significantly more often than those given to nontraumatic words. In that study, however, the stimulus words were not equated for familiarity of usage, and also other variables may have accounted for the results. Thus, Laffal (1955) found that a large proportion of *Ss*' variance in failures to reproduce associations, delayed association time, and other "response faults" were attributable to the interference among mutually competing responses—which he labeled "response entropy."

METHOD

Subjects

The *Ss* were female volunteers from an undergraduate course at the University of Michigan. Thirty-four participated in the main part of the experiment. They were seen singly and were administered the word association test and the recall test. Free associations, reaction times, and recall responses were obtained for all *Ss*. Galvanic skin responses were secured from only 20 *Ss* (the GSR apparatus was inoperative on a number of occasions).

About 4 months after their original testing, 27 *Ss*[3] returned for retesting. These *Ss* again gave their free associations to the original list, and rated each of the stimulus words for emotionality.

Procedure

Each *S* was scheduled for a one-hour individual appointment. When the GSR apparatus[4] was in order, one experimenter (E_1) attached the GSR electrodes to the *S*'s left hand and forearm, fastened her hand into a wooden glove, and placed her arm in a comfortable position on a cushion. While E_1 helped *S* relax, the other experimenter (E_2) adjusted the apparatus and waited for the leveling of the skin resistance. E_1 then gave the following instructions:

The first part of this study has to do with people's reactions to different words. I am going to read you a list of words—one at a time. Each time I say a word, I want you to tell me the first word which comes to mind. So, for example, I might say the word "table," and you might then think of the word "chair." In that case you would immediately say "chair." I want you to do this as quickly as possible, since I am going to time you. Do you have any questions?

The list of 60 words was read to *S*. Reaction times were recorded to tenths of seconds. Meanwhile, E_2 recorded *S*'s GSR deflections which followed the presentation of each stimulus word. After each GSR deflection returned to the base point, E_2 flashed a light concealed from *S*, and E_1 could proceed to read the next stimulus word.

At the end of the 60 words, E_1 asked *S* to recall the associations as follows:

That was fine. Let's take a short break.... Now let's go on to the second part of the task. This time I am going to read you the same list of words. But this time, each time I say a word, I want you to *remember* the word you said the *last* time. So, for example, if I had said "table" and you "chair," this time you would try to remember "chair," and say it as soon as you remember it. Do you understand?

It may happen that you will not be able to remember a word. In that case, I want you to try to think of it and guess as close to the word as you can. If you do guess at a word, just tell me that you are guessing. Also, if you think you have forgotten a word and should remember it later on, be sure to tell me the word immediately.

Following these last instructions, the stimulus list was read once more and recall responses were noted.

The Word List

The word list is reproduced in Table 1. Varieties of word association lists have been used widely, ever since Jung's publication of his technique in 1906. Forty of the words in this list were selected from Hull and Lugoff's (1921) English version of Jung's list. Hull and Lugoff studied the reactions of 100 individuals to each of the 100 words in Jung's list. The stimulus words were compared according to their tendency to elicit "complex signs of emotional disturbance"—such as long reaction time, inability to make any response, misunderstanding of the stimulus word, etc.

The 30 neutral words in this study were selected from the least "disturbing" 50 words of the Hull-Lugoff series. Ten of the emotional words were selected from the top half of the series, with some modifications of verbal form (e.g., from "fearing" to "fear"). Since the other words on the Hull-Lugoff list appeared too neutral, the remaining 20 stimuli were selected a priori to represent objects and activities that might produce some emotion in *S*. As shown in Table 1, the two sets of 30 words were approximately equal with regard to their frequency of usage, as reported by Thorndike and Lorge (1944).

Retest Procedure

About 4 months after the original test, the *Ss* were contacted once again and offered remuneration for

[3] The *Ss* were paid to come to the retest session. Of the seven *Ss* who did not return, four were no longer on campus and the remaining three could not be scheduled for an appointment.

[4] The GSR apparatus was designed and built by D. Rigler (cf. his doctoral dissertation, 1957). A constant current of 40 microamperes was passed through *S*, and changes in resistance were measured. Clay disk electrodes were strapped to *S*'s forearm and hand, which was held immobile in a wooden glove.

TABLE 1
THE LIST OF STIMULUS WORDS AND THE DATA

Stimulus Word	T-L[a] Count	Zero Recall	Mean GSR	Mean RT	Mean ER	Pop. Var.	Retest Var.
1. head[b]	AA	0	33.6	1.5	5.00	14	18
2. *man*	AA	3	20.8	1.5	5.81	16	7
3. frog	25	2	24.7	1.6	2.19	14	12
4. *conduct*	A	1	31.0	2.4	5.40	11	10
5. *truth*	AA	6	33.6	2.1	4.85	9	13
6. fur	A	0	37.4	2.0	3.27	11	12
7. *pity*	A	7	34.1	2.8	6.35	20	13
8. lake	AA	0	15.1	1.7	3.59	12	10
9. *deceive*	33	11	24.9	3.6	5.89	16	18
10. *weak*	A	6	22.8	1.9	5.31	9	5
11. ink	20	1	17.3	1.7	2.19	4	8
12. cold	AA	2	21.7	1.9	3.88	12	14
13. *fear*	AA	18	35.8	3.6	6.78	21	23
14. lamp	A	1	19.2	2.0	1.88	7	6
15. *dead*	AA	4	40.4	2.6	5.98	15	14
16. *power*	AA	7	39.5	2.1	5.44	16	14
17. hay	46	1	14.3	2.6	2.35	12	12
18. *mother*	AA	2	35.7	1.8	6.53	11	8
19. village	AA	0	16.4	1.8	4.37	10	5
20. wild	AA	3	37.7	3.5	5.50	26	20
21. *excel*	11	6	37.1	3.0	4.80	25	19
22. *love*	AA	9	34.6	2.1	6.80	12	11
23. sing	AA	6	26.2	2.0	4.93	18	14
24. *friendly*	AA	7	31.2	2.6	6.34	23	17
25. *strong*	AA	9	18.6	2.2	4.12	12	13
26. painting	45	1	42.3	2.2	5.58	19	16
27. *despise*	26	0	30.8	2.1	6.19	6	6
28. *compete*	11	6	24.4	3.0	5.61	16	16
29. month	AA	0	24.1	1.7	2.58	8	9
30. *bad*	AA	6	13.7	2.0	5.31	8	4
31. cow	A	0	19.3	1.8	1.92	9	15
32. new	AA	3	18.3	1.9	3.92	7	11
33. *angry*	A	14	14.9	2.2	6.80	16	16
34. tree	AA	1	15.8	1.8	2.70	15	15
35. *woman*	AA	0	14.7	1.7	6.00	5	6
36. mountain	AA	0	21.1	1.9	3.12	15	14
37. carrot	8	0	23.3	2.2	2.04	19	14
38. *hate*	A	12	21.4	2.3	6.90	10	13
39. long	AA	4	22.2	2.4	2.66	11	13
40. *ridicule*	9	1	30.8	2.4	6.09	19	15
41. *kiss*	AA	3	58.6	2.7	6.81	16	8
42. house	AA	6	15.4	2.1	4.62	13	4
43. *false*	A	9	22.0	2.0	5.80	12	11
44. window	AA	4	29.5	1.9	2.42	13	16
45. *father*	AA	1	14.6	1.5	6.17	4	4
46. door	AA	1	19.4	1.9	2.24	7	12
47. pencil	40	1	12.8	1.4	1.73	8	16
48. *sin*	A	9	45.1	3.1	6.53	15	18
49. *quarrel*	A	3	13.1	1.9	5.94	12	10
50. hunger	37	1	18.2	1.9	4.89	11	14
51. *marry*	AA	1	50.0	3.3	6.90	18	17
52. voyage	41	1	22.4	1.9	4.66	11	11
53. salt	AA	0	11.8	2.5	2.08	9	9
54. *damage*	32	5	14.4	2.3	4.59	22	17
55. bird	AA	1	21.1	1.8	3.12	16	19
56. *good*	AA	3	13.8	1.6	5.83	5	3
57. book	AA	1	19.3	1.5	4.39	11	16
58. *beating*	5	9	29.9	2.6	6.05	21	20
59. old	AA	7	14.1	1.8	4.55	14	12
60. big	AA	2	14.7	1.5	2.92	12	13

[a] T-L Count refers to the frequency of each word per million words in the English language, as reported by Thorndike and Lorge (1944). AA means more than 100 per million; A means more than 50 per million.

[b] Stimulus words printed in ordinary type were selected by *E*s as "neutral," and italicized words were selected as "emotional."

returning for another appointment. This time *S*s were scheduled in groups ranging in size from 2 to 14. The word list was administered again, with the same instructions as given during the first part of the previous procedure. *S*s wrote their associations into a 60-page booklet, being allowed 8 seconds between the presentation of one stimulus word and the next.

In the second part of the retest period, *S*s rated each of the stimulus words on several seven-point scales, including one which ranged from "emotional" to "unemotional." Each *S* received a 60-page booklet containing the scales on every page. In giving the instructions, the *E* said that, even when it seemed difficult to rate a word on a scale, one had to make the best possible try and assign some rating. Each word was presented for 18 seconds.[5]

Indices

The following indices were constructed as measures of recall, of emotion, of response variability, and reaction time.

Recall. Responses on the recall test were classified into three categories: Total Recall, Zero Recall, Partial Recall. Total Recall refers to all answers which were identical to the *S*'s original association; Zero Recall refers either to no answer at all or to clearly incorrect ones; Partial Recall refers either to incorrect answers similar to the original response (e.g., "cool" after "cold," or "trip" after "journey") or to correct answers which occurred after *S* had already given an erroneous one. On the recall test, 85.8% of the responses showed Total Recall, 11.1% Zero Recall, and 3.0% Partial Recall. All but one *S* forgot at least two associations, but none forgot more than 14.[6]

Emotion. Three criteria of emotion were employed in this study. The first was galvanic skin response. The GSR index consisted of changes in the *S*'s resistance while she gave each word association. All GSR deflections were converted to standard scores as ratios of the *S*'s base resistance, $\Delta R/R$ (see Woodworth & Schlosberg, 1955, p. 157). By this method, valid comparisons could be made among the responses for any given *S*.[7]

The second index consisted of the a priori classifica-

[5] Between the two retest tasks, two rather brief personality tests were administered in connection with some exploratory research.

[6] It is assumed that all failures by *S*s to repeat their initial response, when instructed to remember it on the second stimulus presentation, constituted inadequate retention. When these occurred, *S*s tried hard to recall the missing word; occasionally they would give a delayed correct response, and other times they would manage partial recall.

[7] When GSR scores were averaged across *S*s, an interesting finding emerged. For the 26 stimulus words which were identical in this study and Smith's study (1922), there was a moderately high correlation between the GSRs—rho = .56. After adding four more words, presented in slightly different form in the two studies (e.g., "beating" vs. "beat"), rho increased to .60. Such a correlation between results more than 25 years apart bolsters one's confidence in the potential stability of a GSR index.

tion of the stimulus words by the *E*s as either "emotional" or "neutral." The third index was the emotionality of the stimulus words as rated by the *S*s during the second testing session.

Response variability. Two indices were constructed pertaining to a response-competition explanation of forgetting. One was called Population Variability: the number of different responses by the 34 *S*s given to each stimulus word on the initial association test. The other was Retest Variability: the number of different responses given by the (retested) *S*s to each stimulus word in the two different administrations 4 months apart. In constructing these indices, it was assumed that the greater the number of different responses for a stimulus, the greater would be the potential competition (or interference) among alternative responses.

Reaction time. This variable was measured in tenths of seconds by E_1, from the time he began to say the stimulus word to the beginning of *S*'s response.

HYPOTHESES AND RESULTS

Hypothesis 1

Associations that are subsequently forgotten have higher GSR deflections than those subsequently remembered. This hypothesis was supported. For each of the 20 *S*s for whom GSRs were available, the mean difference between scores on the Zero Recall and the Total Recall associations was computed. Although not all *S*s had higher average deflections during the former responses, the difference was significant at beyond the .05 level ($T = 51$, $N = 20$, two-tailed $p < .05$, by Wilcoxon's matched pair signed-rank test; cf. Wilcoxon, 1949).

Hypothesis 2

Associations to "emotional" words, on the stimulus list selected by the *E*s, show greater forgetting than associations to "neutral" words. This hypothesis received strong confirmation. Comparing each *S*'s number of Zero Recall associations for the emotional and the neutral words, it was found that 31 *S*s forgot more of the former, 3 *S*s forgot an equal number of each, and none forgot more of the latter. (By sign test, this difference is significant far beyond the .001 level.) Of the total of 228 Zero Recall associations (cf. Table 1), more than 78% were responses to stimuli initially classified as "emotional." For example, 18 *S*s—more than half—failed to recall their responses to "fear"; 14 and 9 did not recall their associations to "angry" and "love," respectively; but no one forgot on such neutral words as "head," "fur," "lake."

Hypothesis 3

Associations to words high on emotionality, as rated by the *S*s, show greater forgetting than those low on emotionality. This hypothesis also was confirmed beyond the .001 level of significance, by sign test. Twenty-eight *S*s had more Zero Recall for associations to stimuli that were above the median on emotionality, two for those given to stimuli below the median, while four *S*s forgot an equal number of associations to high and low emotional stimuli.

These findings may appear to demonstrate that the forgetting of word associations was indeed related to emotional factors. One cannot accept this conclusion, however, until one has examined the alternative explanation that the forgetting of such responses is determined by the degree of competition among alternative responses. Laffal's (1955) finding that "response faults" on word association tests, long considered to be indicators of emotional disturbance, are highly correlated with "response entropy," referring to the uncertainty that any single response is connected with the stimulus, is illustrative of this line of explanation. Our index of Zero Recall is one kind of response fault, although this index has been little used in previous studies.

Following this line of reasoning, we examined the relation between forgetting (Zero Recall) and the two indices of response competition (Population and Retest Variability). The relations were almost as high as in Hypotheses 2 and 3, and also significant beyond the .001 level. Twenty-six *S*s forgot more associations to stimuli with high Population Variability, and only three did the opposite. Twenty-four *S*s showed higher Zero Recall for responses to stimuli with high Retest Variability, and only six showed lower forgetting for such responses.

Thus, it was found that forgetting was almost equally related to the indices of emotional significance and those of response interference, although perhaps slightly more to the former indices. This finding led to a further exploration among the different response indices. It was asked: Can the forgetting of word associations be accounted for in terms of *either* emotional inhibition or response com-

TABLE 2
INTERCORRELATIONS AMONG THE RESPONSE VARIABLES[a]

Variable	Mean GSR	Mean RT	E-N[b]	Mean Emot.	Pop. Var.	Ret. Var.
Zero Recall	.16	.49	.55	.50	.38	.28
Mean GSR	—	.50	.28	.43	.45	.40
Mean Reaction Time		—	.36	.37	.58	.50
"Emotional–Neutral"			—	.76	.16	−.04
Mean Emotionality				—	.31	.09
Population Variability					—	.71
Retest Variability						—

[a] Correlations were by Pearson *r*, except for the use of point biserials for all correlations with the dichotomous "emotional–neutral" variable.

[b] A positive correlation with the E-N variable means a positive relationship with the emotional end.

petition, or did *each* variable account for different parts of the outcome?

Whereas until now the findings have pertained to the 34 *individuals* in the sample, the following analysis was performed on the responses to the set of 60 *words* in the stimulus list. Since the distributions of most of the variables were rather skewed, the scores for each variable were transformed to a normal distribution (cf. Edwards, 1954, pp. 107–111).

Correlations among the variables are shown in Table 2.[8] It is evident that Zero Recall was correlated with all variables except GSR.

To explore these relationships further, some partial correlations were computed between Zero Recall and other variables, removing the effects of Population Variability and Retest Variability. Doing this, it was found that the relation between Zero Recall and Mean Emotionality, for example, was little changed. The partial correlations were .44 and .50, removing Population and Retest Variability, respectively.

This result demonstrated the apparent existence of two independent determinants of forgetting. It led to a more systematic examination of the correlational matrix—namely, a principal components factor analysis (cf. Hotelling, 1933). In the factor analysis, two

[8] The following correlational analysis does not assume that the 60 words were a random sample of all possible words, since the words were selected to represent different points on an assumed continuum of emotionality. For this reason, the *statistical* significance of the correlations cannot be assessed. However, the correlations do have *substantive* significance, since they show the amount of variance in Zero Recall attributable to each variable.

TABLE 3
FACTOR LOADINGS OF RESPONSE VARIABLES

Variable	Principal Axes		Rotated Axes	
	I	II	I′	II′
Zero Recall	.63	−.19	.59	.29
Mean GSR	.58	.15	.33	.50
Mean Reaction Time	.77	.12	.50	.60
"Emotional–Neutral"	.62	−.62	.86	−.03
Mean Emotionality	.75	−.46	.86	.18
Population Variability	.70	.46	.20	.81
Retest Variability	.55	.64	−.04	.84

main dimensions appear. Table 3 shows the results for Principal Axes I and II, and for the Rotated Axes I′ and II′. Factor I′ seems to represent the emotionality dimension, while Factor II′ shows response variability. The remaining three variables are located between these two dimensions. GSR and Reaction Time fall more toward the variability factor, while Zero Recall is clearly biased toward emotionality.

Only two important factors were found. In locating the distribution of the remaining variance, a third factor was found most loaded with GSR (−.44) and Zero Recall (+.39), while all other loadings remained small.

It seems clear that the emotionality and the variability measures were relatively independent, while both were related to forgetting. What remains unknown is the nature of the relationships. The factor analysis does not justify the conclusion that forgetting is determined more by emotional factors, since it demonstrates only statistical and not causal connections. Yet statistical relations are a necessary condition for causal ones.

The findings lead one to conclude that the forgetting of word associations was a function of two independent factors—the emotion-arousing potentiality of the stimulus words and their tendency to evoke mutually competing responses. This conclusion seems similar to one reported by DeLucia and Stagner (1954) from a study of word recognition. They noted that word recognition time was clearly affected by " . . . two sets of determinants: frequency of usage and emotion-arousing value" (p. 309).

Before turning to a discussion of the results, some findings about frequency of word usage are presented. The frequency of usage of the "emotional" and "neutral" *stimulus words*

was equal with reference to the Thorndike-Lorge word count, as mentioned earlier. When the varying frequency of these words was related to their tendency to evoke Zero Recall associations, no difference was found between high and low frequency stimuli (by median test, $\chi^2 = 0.27$; not significant).

Comparing the word frequencies of the Total Recall *associations* with those of the Zero Recall ones, it was found that the frequencies of the former were higher than those of the latter. It is possible, then, that the relatively greater unfamiliarity of certain associations may have played a part in *S*s' forgetting. Nevertheless, the difference between the two kinds of associations was substantively small: e.g., while 83% of the remembered associations occurred 50 or more times per million words in the Thorndike-Lorge count, over 70% of the forgotten associations fell in this same category. Thus, the frequency in the English language of the stimulus and response words did not appear to have an important effect on the occurrence of forgetting.

DISCUSSION

Three different indices of the emotional import of the stimulus words were found to be significantly related to *S*s' failure to recall their associations. Such forgetting was unrelated to the word frequency of the stimuli, and had little relation to the word frequency of the responses. What is the meaning of these results?

The findings have been considered in terms of two different explanations, which both focus on the strength of the association between stimulus and response. One is based on an hypothesis of response variability, which asserts that probability of recall depends on the slope of the response hierarchy, as inferred from the frequencies with which different responses have been associated to the stimulus in the past (cf. Osgood, 1953, pp. 550–551). The other explanation invokes the hypothesis that the more emotional the stimulus, the weaker the connection between it and the specific response by *S* in the association test. The validity of this hypothesis depends on the presumed existence of an "emotional inhibition" process, operating at the moment when *S* makes the association.

The hypothesis of response variability bases its inferences on general population parameters, which suggest the probability of the occurrence and re-occurence of various stimulus-response connections. The hypothesis of emotional inhibition does not contradict the validity of that approach, but focuses on more specific blocking or failures to recall. Since specific predictions are usually more erroneous than general ones, it is not surprising that little evidence for the validity of the emotional inhibition hypothesis has yet been found.

The present evidence leads us to propose that the two different explanations are supplementary, and not contradictory. Let us briefly review the data.

The finding that GSRs on the first presentation were higher for subsequently forgotten than for subsequently remembered associations is not central to this discussion. Actually GSR, assumed to be an index of emotional activation during the associative process, bore a higher relation to response variability than to stimulus emotionality.

On the other hand, as seen in the correlational analysis, the stimuli regarded as emotional by *E*s and *S*s evoked greater forgetting of associations than did those regarded as neutral. Furthermore, Zero Recall was found to be independently correlated with these indices of emotionality and with those of response variability—the former correlations being higher than the latter. Although the nature of the data prevents the assessment of their statistical significance, this exploratory analysis does raise certain questions for future studies.

First, to what extent is forgetting determined by emotionality? The results of this study merely demonstrate certain correlations between these variables. Yet, in the writers' opinion, they lend some support to the hypothesis that emotional factors can determine forgetting. It seems unlikely that an "emotionless" theory of association can explain these findings.

Second, what is the relation between the emotional significance of stimulus words and the variability within response hierarchies? The present study found relatively low, but positive, correlations between these sets of

variables. Yet it remains possible that emotional words in general tend to elicit more competition among alternative responses than do unemotional ones. And if this were so, why? The answer to such a question does not necessarily lie within the frame of a psychoanalytic repression model. On the other hand, it does not appear to be provided by more systematically supported theories of memory and forgetting.

SUMMARY

The study aimed to demonstrate that the forgetting of word associations is related positively to their emotional significance.

Thirty-four *Ss* were administered a list of stimulus words, containing half "emotional" and half "neutral" words. *Ss* gave free associations to the words, while measures of galvanic skin response and reaction time were obtained. Immediately thereafter, the words were presented again in the same order, and *Ss* tried to recall their previous associations. Four months later, *Ss* returned and again gave free associations to the same stimuli, and also rated the emotionality of each stimulus word.

It was found that *Ss* showed significantly higher GSRs while giving associations which were later forgotten than while giving those which were later remembered. Also, stimulus words rated highly emotional by (*a*) the *Es*, and (*b*) the *Ss*, were far more likely to elicit associations which were later forgotten than were those stimuli rated low on emotionality.

It was also found that stimulus words which evoked high variability in responses (*a*) among different *Ss*, and (*b*) between tests for the same *Ss*, were more likely to have their associations forgotten than those which had low variability in responses.

In a factor analysis of the correlations among different variables, it was discovered that emotionality and response variability were independently related to forgetting. It was proposed that the forgetting of word associations is a function of both emotional and nonemotional determinants.

REFERENCES

BUGELSKI, B. R. *The psychology of learning.* New York: Holt, 1956.

DELUCIA, J. J., & STAGNER, R. Emotional vs. frequency factors in word-recognition and association time. *J. Pers.*, 1954, **22,** 299–309.

EDWARDS, A. L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954.

GRUMMON, D. L., & BUTLER, J. M. Another failure to replicate Keet's study, "Two verbal techniques in a miniature counseling situation." *J. abnorm. soc. Psychol.*, 1953, **48,** 597.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. *J. educ. Psychol.*, 1933, **24,** 417–441, 498–520.

HULL, C. L., & LUGOFF, L. S. Complex signs in diagnostic free association. *J. exp. Psychol.*, 1921, **4,** 111–136.

JUNG, C. G. *Diagnostische Associationsstudien.* Part I. Leipzig: Barth, 1906.

KEET, C. D. Two verbal techniques in a miniature counseling situation. *Psychol. Monogr.*, 1948, **62** (7, Whole No. 294).

LAFFAL, J. Response faults in word association as a function of response entropy. *J. abnorm. soc. Psychol.*, 1955, **50,** 265–270.

MELTZER, H. The present status of experimental studies on the relationship of feeling to memory. *Psychol. Rev.*, 1930, **27,** 124–139.

MERRILL, R. M. On Keet's study, "Two verbal techniques in a miniature counseling situation." *J. abnorm. soc. Psychol.*, 1952, **47,** 722.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford, 1953.

RAPAPORT, D. *Emotions and memory.* Baltimore: Williams & Wilkins, 1942. (Reprinted: New York: International Univer. Press, 1950.)

RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing.* Vol. II. Chicago: Year Book Publishers, 1946.

RIGLER, D. Some determinants of therapist behavior. Unpublished doctoral dissertation, University of Michigan, 1957.

SMITH, W. W. *The measurement of emotion.* New York: Harcourt, Brace, 1922.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Bureau of Publications, 1944.

UNDERWOOD, B. Interference and forgetting. *Psychol. Rev.*, 1957, **64,** 49–60.

WILCOXON, F. *Some rapid approximate statistical procedures.* New York: American Cyanamid Company, 1949.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* (Rev. ed.) New York: Holt, 1955.

ZELLER, A. F. An experimental analogue of repression: I. Historical summary. *Psychol. Bull.*, 1950, **47,** 39–51.

(Received December 21, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 106–113

RESPONSIVENESS OF CHRONIC SCHIZOPHRENICS AND NORMALS TO SOCIAL AND NONSOCIAL FEEDBACK[1]

WALTER J. JOHANNSEN[2]

Veterans Administration Hospital, Wood, Wisconsin

DURING the last 2 decades, psychological research interest in schizophrenia has more and more frequently involved a study of social factors related to the syndrome. Much of this interest has stemmed from theoretical issues raised by Cameron and by Sullivan. Both of these theorists viewed schizophrenia as a disorder arising out of faulty social adjustment. This point of view has received considerable support. Studies dealing with this "social dysfunction hypothesis" (e.g., Affleck, 1954; Davis & Harrington, 1957; Flavell, 1956; Whiteman, 1954) have shown that the functions of motivation, concept formation, and learning in schizophrenia are disrupted by social features of test stimuli or presentation.

Despite the growing mass of data accumulated in support of the Cameron-Sullivan theories, a contention basic to both their arguments demands resolution. This fundamental assertion holds that schizophrenics display impairment in the ability to codify information about their behavior when that information derives from an external, social source. This statement underlies both Sullivan's notions of consensual validation and Cameron's role playing theory. Establishing a schizophrenic deficit in the ability to respond to social feedback[3] would be consistent with the Cameron-Sullivan position and would strongly affirm the utility of the social dysfunction hypothesis.

The following study was designed to test four hypotheses:

1. Schizophrenics are significantly poorer than normals in modifying behavior on a learning task where solution requires reception and codification of social feedback.

2. Among schizophrenics, impairment is significantly less when the source of feedback is nonsocial than when the source is social. This prediction follows from the contention of the social theorists that social interaction is the disruptive element in schizophrenic performance. Where interaction is reduced, performance approaches the level established by normal controls.

3. Paranoid schizophrenics function more adequately on this task than nonparanoid schizophrenics but not as well as normals. This prediction derives from Sullivan's assertion that paranoid schizophrenics' communicative impairment is less severe and genetically later in origin than that of the other subgroups.

4. Normal controls, who do not show social maladjustment tendencies peculiar to schizophrenics, perform equally well under social and nonsocial feedback conditions.

METHOD

Subjects

A total of 125 subjects were tested. Twenty-four nonparanoid schizophrenics served under social feedback conditions and 24 under nonsocial feedback conditions. The same number of normals served in equivalent feedback conditions. Fifteen paranoid schizophrenics served in the social feedback condition. 14 in the nonsocial condition. The number of subjects chosen within each diagnostic category was established in order to maintain approximate proportionality of sample size to the size of the parent population.

Schizophrenic subjects were selected from among the patients of the Tomah Veterans Administration hospital. Normal controls were chosen from the patients of the general medical service of the Wood Veterans Administration center.

Schizophrenic subjects were classified on the basis of current psychiatric diagnosis. This population was then screened to eliminate individuals deemed unsuitable for the study. Patients were excluded for the

[1] This research is abridged from a portion of a doctoral dissertation submitted to the University of Wisconsin.

[2] The author is indebted to his major professor, Richard M. Lundy, and to L. Benjamin Wyckoff, Samuel H. Friedman, and Carl R. Rogers for assistance and criticism during various stages in the completion of this project.

[3] "Feedback" is taken to mean the signaling of information about performance to the subject. "Social feedback" is such information, the immediate source of which is another human individual. "Nonsocial feedback" is information whose source is other than another human individual.

following reasons: age under 20 or over 59, total hospitalization of less than a full year, psychosurgery, and insulin or electroshock therapy during the preceding month. Patients whose case histories indicated a change in diagnosis from paranoid to nonparanoid schizophrenia were dropped from consideration. Finally, in order to avoid confounding of functional psychosis with established organic conditions, patients bearing complicating neurological diagnoses were eliminated. A total of 271 nonparanoid schizophrenics and 143 paranoid schizophrenics remained available.

From this subject pool, 48 nonparanoid schizophrenics were randomly selected and assigned to one or the other of the two feedback conditions. Two paranoid schizophrenic groups were then selected, equated with the nonparanoid schizophrenics on age, years of hospitalization, and educational attainment. Normal subjects were selected to provide groups equated with the schizophrenic groups on age and length of education. Adequate control of the hospitalization variable proved to be impossible. To maintain consistency of selection, patients bearing neurological or psychiatric diagnoses were excluded from the control group. A summary of the pertinent actuarial characteristics is presented in Table 1.

Task and Apparatus

The task presented to all subjects was the learning of a repetitive double alternation pattern (RRLLRRLL). Two hundred trials were given all subjects. Each trial required the subject to respond by pressing one of two buttons. This type of task appeared ideal for a study of this nature, since conditions could be manipulated so that the solution required codification of externally manipulated feedback information. The subject could not evaluate the correctness of response without attention to the feedback signal.

Subjects were seated at a desk with an unpainted 6″ × 10″ wooden panel placed flat on the desk surface approximately 10″ in front of them. On this response board were mounted two doorbell buttons, one 2″ to the right and the other 2″ to the left of center.

In the social feedback condition, the experimenter sat beside the desk and at a 90 degree angle to the subject. Response buttons were wired to two 25-watt light bulbs mounted on a panel placed inconspicuously on the floor next to the experimenter. These lights served merely to signal to the experimenter which of the two buttons had been pressed.

In the nonsocial feedback condition, the response board was placed as in the social feedback condition. A panel, 24″ × 40″, was mounted vertically behind it. This feedback panel was so situated that the subject could maintain his gaze on it while responding. Centered near the top were two 25-watt bulbs separated vertically by 5″. To increase discriminability between the feedback signals, the top bulb was red and the bottom bulb white.

These lights were the immediate source of feedback and were controlled by the experimenter, seated in an adjoining room. The subject's response buttons were wired to two bulbs mounted on a panel situated in front of the experimenter; from this position he signaled information to the subject's feedback panel.

Procedure and Instructions

Subjects were run individually. Prior to the experimental session, each subject was engaged in conversation for about 5 minutes in order to secure rapport. Instructions then given were designed to avoid the establishment of an atmosphere related to personal achievement. All subjects were initially told the following:

> All right, we're going to play a game now. You see these buttons in front of you? All I want you to do is press them, one at a time, and in any order you want. Before we get into the game, why don't you go ahead and press them a few dozen times, just to get the idea of it. Remember now—one at a time and in any order you want. Do it slowly.

Subject was then encouraged to respond 40 times before the experimenter called a halt. No feedback was given during this sequence. The procedure was meant to accustom the subject to the task, to establish a slow speed of response, and to provide a means of noting possible stable response tendencies. This record of response tendencies was used in data analysis.

After completion of the free responding sequence all subjects were given the following basic instructions:

> Good. You've got the idea. Now in this game we're going to play, sometimes one of these buttons is going to be the correct one to press and sometimes the other button is going to be the correct one to press. This is the way we're going to do it.

At this point, subjects serving in the social feedback condition were told:

> From now on I will tell you after each time whether you pressed the right one or the wrong one. If you pressed the correct button, I'll say "right"; if you pressed the wrong one I'll say

TABLE 1

ACTUARIAL CHARACTERISTICS OF DIAGNOSTIC GROUPS

Group	Age (years)		Hospitalization (years)		Education (years)		Vocabulary Score	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Nonparanoid Schizophrenics	38.29	6.38	9.67	3.68	10.58	2.69	25.54	5.85
Paranoid Schizophrenics	38.03	6.87	7.08	2.94	10.73	1.91	24.97	7.44
Normals	38.37	6.46	—	—	9.65	2.90	25.24	7.75

"wrong." Each time you press I will tell you whether you were right or wrong. Okay? Remember now, sometimes one of these buttons is going to be the right one to press and sometimes the other button is going to be the right one to press. Try to get as many correct as you can. Okay? Go ahead.

Throughout the experimental series, the experimenter, seated close by the subject and in his full view, responded either "right" or "wrong" in a level tone of voice. Responses were recorded by the experimenter.

In the nonsocial feedback condition, subjects were given these instructions:

I'm going to leave you alone now. From now on, these lights will tell you after each time whether you pressed the right one or the wrong one. If you pressed the correct button, this top light, the red one, will go on. But if you press the wrong button, this light, the bottom one, will go on. Each time you press one of these buttons one of these lights will go on. It will tell you whether you were right [experimenter points to top light] or whether you were wrong [points to bottom light]. Remember now, sometimes one of these buttons is going to be the correct one to press, and sometimes the other is going to be the correct one to press. Try to get as many correct as you can. Okay? [experimenter, leaving the room] Go ahead.

Subjects who were unable to follow directions were immediately eliminated from the study. In addition, a few subjects refused to participate. A total of 16 subjects were dropped for these reasons (untestable: 8 nonparanoid schizophrenics, 2 paranoid schizophrenics, 0 normals; refused: 1 nonparanoid schizophrenic, 5 paranoid schizophrenics, 0 normals). It is unlikely that these factors had an appreciable biasing effect on the results obtained, since the frequencies involved were small. Substitutions were made on the basis of similarity in pertinent actuarial characteristics.

After the completion of 200 trials, the experimenter returned to the subject and, to establish validity of the data obtained, asked subject which of the two feedback lights signaled a correct response. Then, to obtain a control for intelligence, all subjects were administered the vocabulary scale of the Shipley-Hartford intelligence test. Since research has frequently shown vocabulary is little affected by psychosis (Babcock, 1930; Kendig & Richmond, 1940; Shakow, 1946) this test seemed to offer an adequate estimate of premorbid intelligence.

RESULTS

Evaluation of Sources of Bias

Prior to analysis of the experimental data, the contribution of extraneous factors to the obtained results was estimated. As an initial step, analyses of variance were performed to determine whether groups were adequately equated on the actuarial variables of age, education, and intelligence. None of the resulting F ratios achieved significance; hence, it may be assumed that these three variables did not contribute significantly to the experimental results.

Since it was impossible to equate normals and schizophrenic groups on total length of hospitalization, it was necessary to determine whether this factor contributed a bias to the experimental results. Correlation analysis demonstrated that the degree of relationship between length of hospitalization and task performance was negligible. Thus, group differences in length of hospitalization did not appear to affect significantly any differences in experimental performance.

Another source of bias was the effect of tranquilizing medication on task score. Since nearly 80% of the schizophrenic subjects were receiving such treatment, the effect of tranquilizers on performance required evaluation. Those subjects not receiving drugs were matched with subjects within the same experimental cell who were receiving medication and whose actuarial characteristics were most similar to theirs. Within each cell, these tranquilizer and nontranquilizer groups were dichotomized into frequencies scoring above and below the mean task score for that cell. If medication affected task performance, differences would be expected between the tranquilizer and nontranquilizer groups in respect to distribution about the cell mean. Frequencies were accumulated from all six cells, and the resulting 2×2 contingency table compared by means of the Yates-corrected chi square test. The obtained value of 1.02 falls far short of significance. It, thus, appears that presence of tranquilizing medication did not act significantly to bias the results of the experiment.

One other source of bias required evaluation. It was thought that some subjects would approach the task with stable response pattern sets. A bias toward an RRLL pattern would tend to produce a higher score than an RLRL bias. Inspection of the free responding series data indicated, however, a virtual identity of group means in regard to the number of alternations. Preliminary response sets contributed little bias to the experimental results.

Analysis of Experimental Results

The mean number correct responses was calculated for each experimental condition,

and the results are presented graphically in Figure 1. These data reveal a striking interaction. The normal controls show a markedly better mean performance under social feedback conditions (161.25) than under nonsocial feedback conditions (129.00). This pattern is reversed for the nonparanoid schizophrenics who perform more adequately under nonsocial feedback conditions (127.54) than under social feedback conditions (108.00). Paranoid schizophrenics show an equivalence of performance under social (130.47) and nonsocial (123.64) conditions. It is noteworthy that performance of the three diagnostic groups is almost identical under nonsocial feedback conditions.

Extreme skewedness of cell distributions made obligatory the use of different, relatively independent techniques for evaluating the statistical significance of these results. Since Norton (1953) demonstrated that the *F* test is highly insensitive to deviations from normal distribution, the analysis of variance was used as the basic test of the data, with significance level adjusted to guard against unwarranted rejection of the null hypothesis. In this analysis, one score was randomly eliminated from the paranoid schizophrenic social feedback condition in order to achieve proportionality of cell frequencies. The summary of the analysis of variance is contained in Table 2.

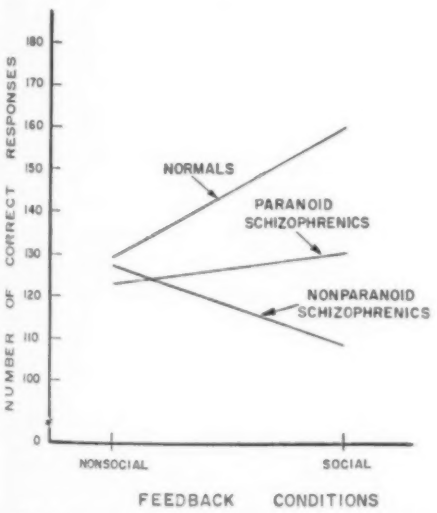

FIG. 1. Mean number correct responses for diagnostic groups as a function of feedback condition.

The results indicate that significant differences exist among diagnostic groups with respect to their responsiveness to feedback, with normals most adequate, nonparanoid schizophrenics least adequate, and paranoid schizophrenics performing in an intermediate position. This result is in accord with the hypothesis.

The significant group × feedback interaction indicates that the three diagnostic groups perform in a different fashion with respect to the nature of the feedback source. This result generally accords with the hypothesis, except that the differentially better performance of normals under social feedback conditions was unanticipated.

Two chi square analyses were then performed to supplement the analysis of variance. The results are summarized in Table 3. In the first of these analyses an overall mean (number of correct responses) was obtained, pooling the data obtained from all subjects. Each cell was then dichotomized with respect to this mean. The frequencies of subjects falling above and below this value were obtained. These data were then subjected to various comparisons.

It is apparent from this analysis that conclusions derived from the analysis of variance are substantiated. Specifically, the three diagnostic groups show significant differences in performance under social feedback conditions but not under nonsocial feedback conditions. Nonparanoid schizophrenics perform significantly better under nonsocial than under social feedback conditions; however, the reverse is true for normals. Paranoid schizophrenics perform equivalently under

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE FOR NUMBER OF CORRECT RESPONSES AS A FUNCTION OF GROUP AND FEEDBACK CONDITION

Source	*df*	*MS*	*F*
Diagnostic Groups	2	9912.52	6.52*
Feedback	1	1329.24	—
Groups × Feedback	2	7300.03	4.79*
Error	118	1522.90	
Total	123		

* Significant beyond the .025 level.

TABLE 3

SUMMARY OF CHI SQUARE ANALYSES COMPARING FREQUENCIES OF SUBJECTS ABOVE AND BELOW OVERALL MEAN (COLUMN A) AND FREQUENCIES OF SOLVERS AND NONSOLVERS (COLUMN B)

Comparison	Chi square (A)	Chi square (B)
Overall cells	22.33**	25.43**
Nonparanoid Schizophrenics		
Social vs. Nonsocial	3.99*	5.73*
Paranoid Schizophrenics		
Social vs. Nonsocial	.01	.01
Normals		
Social vs. Nonsocial	6.84**	10.10**
Social Condition		
All groups	19.22**	21.08**
Nonsocial Condition		
All groups	.18	.80

* Significant beyond the .05 level.
** Significant beyond the .01 level.

either condition. Thus, all hypotheses were again borne out with the single exception of that asserting a similar mode of responding for normals irrespective of feedback source.

A second set of chi square comparisons was then performed. Results are summarized in Column B of Table 3. This analysis was designed to determine whether the number of subjects achieving solution differed from cell to cell. The criterion of correct solution was defined as errorless performance during the terminal 40 trials. Inspection of results shows that this analysis is entirely consistent with the previous analyses. Conclusions based on analysis of the response data apply with equal force here.

One final matter requires brief consideration. An inspection of protocols revealed a large number of consecutive RL or LR alternations among nonparanoid schizophrenic subjects. This form of stereotypy was relatively less frequent among paranoid schizophrenics and even less in evidence among normals. The mean frequencies of RL or LR alternations within each condition are reported in Table 4.

While errorless performance on this task entails 100 such alternations, scores much in excess of this value may be considered indicative of inefficient perseverative behavior. An analysis of variance was performed on alternation scores to determine whether significant differences existed among groups. As shown

TABLE 4

MEAN NUMBER SINGLE ALTERNATIONS FOR EACH CONDITION

Condition	No. Alternations
Nonparanoid Schizophrenics	
Social	151.96
Nonsocial	115.96
Paranoid Schizophrenics	
Social	119.35
Nonsocial	127.07
Normals	
Social	106.00
Nonsocial	112.75

TABLE 5

ANALYSIS OF VARIANCE FOR THE NUMBER OF ALTERNATIONS IN 200 TRIALS

Source	*df*	*MS*	*F*
Diagnostic Groups	2	7266.11	5.48*
Feedback	1	5597.96	4.22
Groups × Feedback	2	5476.69	4.13*
Error	118	1325.47	
Total	123		

* Significant beyond the .05 level.

in Table 5, the three diagnostic groups differed significantly in the mean number of alternations, with nonparanoid schizophrenics showing most and normals least perseveration. The significant interaction term indicates that the three groups responded differently to the feedback conditions. Nonparanoid schizophrenics gave more alternations under social feedback conditions than under nonsocial feedback conditions; paranoid schizophrenics and normals alternated more frequently under nonsocial than social feedback conditions.

DISCUSSION

The hypotheses put forth in this study were in general supported. In a learning task with adequate performance dependent on codification of feedback derived from an outside source, schizophrenics achieved significantly poorer scores than normals. Paranoid schizophrenics proved to be less impaired than nonparanoid schizophrenics. As predicted, schizophrenics performed as well as normals under nonsocial feedback conditions, but were significantly poorer under social feedback conditions. It was also demonstrated, in accordance with the hypothesis, that paranoid schizophrenics

achieve a level of performance intermediate between that of nonparanoid schizophrenics and normals. The single hypothesis which was not borne out dealt with an anticipated equivalence of response level of normals under social and nonsocial feedback conditions. Normals actually showed significantly better performance under social than nonsocial feedback conditions.

These results have implications for several areas of research. On the broadest level of generality, the finding that the schizophrenic is able to function on a normal level in the presence of nonsocial feedback bears directly on the problem of schizophrenic intellectual deficit. The present data lend considerable support to those theorists who contend that this deficit is a matter of ability rather than capacity and that, as a consequence, it is reversible (Hunt & Cofer, 1944; Kendig & Richmond, 1940; Shakow, 1946). Conversely, there is no support for the position that schizophrenic deficit is similar to brain damage, entailing loss of capacity to learn or conceptualize.

From the viewpoint of intellectual function, it is noteworthy that schizophrenic deficit in the present study was most severe under social feedback conditions. It is well known that attitudinal and motivational factors are sensitive to interpersonal influence, and clinical judgment suggests that schizophrenics are particularly attuned to nuances of social interaction. As Wegrocki (1940) says, "The test situation need not represent, then, for the schizophrene, simply a trial ground of his intellectual prowess, but may be a situation fraught with affective innuendoes of which the experimenter is totally unaware" (p. 63).

It is tempting to speculate whether the traditional schizophrenic intellectual dysfunction is not merely attributable to disruption deriving from the presence of an examiner in the testing room. Would the performance of schizophrenics on circular mazes or serial or motor learning tasks show the usual decrement if testing were conducted with the experimenter absent from the room? In support of this conjecture it should be noted that where explanations of deficit are attempted, these attempts are usually couched in motivational terms (Malmo, Shagass, & Smith, 1951; Rickers-Ovsiankina, 1937; Ricksher, 1909; Wegrocki, 1940). Moreover, where intellectual decrement is most severe, the tasks used are most often those requiring prolonged attention and interest (Shakow, 1946).

The nature of schizophrenic functioning under social feedback conditions introduces a further, related question: What aspect of the social feedback situation led to the extreme degree of disruption noted? One fruitful avenue of speculation is suggested by Garmezy's study (1952). He noted that the use of negative evaluation after incorrect response in a discrimination task led to increased adequacy of performance in normals but no change in response among schizophrenics. He argued that a fear of censure was operating differentially to inhibit schizophrenic performance relative to normals. This interpretation is consistent with the Cameron position and observations made during the administration of the present study. Cameron states (1947, p. 486) that schizophrenia develops out of a reaction sensitivity to instances of failure.[4]

Mednick's recent formulation of the development of schizophrenia (1958) offers a convenient explanation for the observed phenomena when a few additional assumptions are made. Mednick contends that schizophrenia is a response to anxiety. Anxiety is conceptualized as a drive state and incipient schizophrenics as individuals continually functioning under conditions of anxiety. High drive acts to increase the response strength of any habit that may be aroused in a given situation as well as producing heightened stimulus generalization (a concept similar to Cameron's "overinclusion"). A process is then initiated in which anxiety and generalization mutually augment each other. The pattern eventuates in responses which are highly irrelevant and inappropriate. The ensuing process of chronicity which takes place is described as one which

[4] Mention must be made of an alternative explanation deriving from an artifact of the experimental situation, namely, the confounding of feedback with sense modality. That is, social feedback was presented in an auditory fashion and nonsocial feedback visually. Differential sensitivity of schizophrenics to one or the other mode of presentation may be sufficient to explain the results. However, since the literature does not demonstrate differential sensory sensitivity in schizophrenics, this explanation is not supported.

involves increasing use of irrelevancy as a method of escaping anxiety.

In explaining the current results, it is assumed that social feedback arouses anticipation of censure and consequently increased drive level among schizophrenics. Nonsocial feedback conditions, however, involve no such change in drive level. It is also assumed that normals respond to increased drive in problem solving situations in a manner defined as adaptive, but that schizophrenics respond in a manner that involves retreat into irrelevancies. That is, confronted with a fear of censure, the normal responds with an attempt to achieve an adequate solution by concentrating, focusing attention, formulating hypotheses, etc. The schizophrenic, especially the nonparanoid schizophrenic, responds by withdrawing from the requirements of the task, i.e., random responding or stereotypy.

The equivalence of paranoid schizophrenic performance under the two feedback conditions is in some ways puzzling. One possible explanation is that paranoid schizophrenics share in the basic interpersonal difficulty of schizophrenics in general, operative under social feedback conditions, but that their ability to profit from feedback is less impaired than is the case among other schizophrenics. While this explanation is plausible, there are alternatives. For instance, it may be that the paranoid schizophrenic relationship to others is one that involves a dehumanization of others; thus, perception of a social feedback source is not different for them from their perception of any other feedback source. Another possibility is that the interpersonal disorder of the paranoid schizophrenic is of the same nature and severity as that of other schizophrenics, but that a degree of attentiveness to social stimuli is maintained which is not matched by other schizophrenics. The present data do not allow a choice among these alternatives; further research is needed in this area to clarify the point raised.

The nonparanoid schizophrenic predilection for the single alternation tendency deserves consideration. Statistical analysis which demonstrated differences among diagnostic groups cannot indicate whether the prevalence of single alternations prevents the achievement of solution, or whether inability to arrive at solution results in prevalence of single alternations. However, the subject who cannot arrive at a solution has few alternatives available to him. He may withdraw entirely from the task by refusing to respond or insisting on the meaninglessness of it. If, however, he does respond, he may continually modify his preliminary response disposition in accordance with feedback provided him in order to achieve a solution. On the other hand, he may respond in a random pattern or in a stereotyped manner. Several studies (Fey, 1951; Kerner, 1956; Peters, 1953; Wulfeck, 1941) have noted sterotypy among schizophrenics under conditions of stress. These findings are consistent with the interpretation that nonparanoid schizophrenics, functioning in an atmosphere involving fear of censure, tend to respond in a stereotyped fashion. This mode of response interferes with the development of more adaptive and successful patterns of response.

SUMMARY

This study was designed to test the hypothesis that schizophrenics' ability to learn under conditions of social feedback is impaired. Subjects were given the task of learning a repeated double alternation pattern under either social feedback (experimenter present in the experimental room, signaling feedback verbally) or nonsocial feedback (experimenter absent from the room; feedback signaled by lights). A total of 125 subjects (48 nonparanoid schizophrenics, 29 paranoid schizophrenics, 48 normals) were tested. The groups were equated for age, intelligence, and education.

Results were generally as predicted. Normals showed the highest level of attainment, nonparanoid schizophrenics the lowest, and paranoid schizophrenics an intermediate degree. The most noteworthy finding was that normals performed significantly better under social feedback conditions than under nonsocial feedback conditions while nonparanoid schizophrenics were significantly poorer under social than under nonsocial feedback conditions. All three diagnostic groups performed equally well under nonsocial feedback conditions.

Discussion centered about the question of nonparanoid schizophrenic performance under the two feedback conditions and its difference

from the normals' mode of responding. An explanation was based on an hypothesis dealing with differential response to a fear of censure.

REFERENCES

AFFLECK, D. C. The effect of interpersonal situations on conceptual performance in schizophrenia. *Dissertation Abstr.*, 1954, **14**, 2399. (Abstract)

BABCOCK, HARRIET. An experiment in the measurement of mental deterioration. *Arch. Psychol.*, 1930, No. 117.

CAMERON, N. *The psychology of behavior disorders: A biosocial interpretation.* Cambridge, Mass.: Houghton Mifflin, 1947.

DAVIS, R. H., & HARRINGTON, R. W. The effect of stimulus class on the problem-solving behavior of schizophrenics and normals. *J. abnorm. soc. Psychol.*, 1957, **54**, 126–128.

FEY, ELIZABETH T. The performance of young schizophrenics and young normals on the Wisconsin Card Sorting Test. *J. consult. Psychol.*, 1951, **15**, 311–319.

FLAVELL, J. H. Abstract thinking and social behavior in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **52**, 208–211.

GARMEZY, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of reward and punishment. *J. Pers.*, 1952, **20**, 253–276.

HUNT, J. MCV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944. Pp. 971–1032.

KENDIG, ISABELLE V., & RICHMOND, W. V. *Psychological studies in dementia praecox.* Ann Arbor, Mich.: Edwards, 1940.

KERNER, O. J. B. Stress, fantasy, and schizophrenia: A study of the adaptive process. *Genet. psychol. Monogr.*, 1956, **53**, 189–281.

MALMO, R. B., SHAGASS, C., & SMITH, A. A. Responsiveness in chronic schizophrenia. *J. Pers.*, 1951, **19**, 359–375.

MEDNICK, S. A. A learning theory approach to research in schizophrenia. *Psychol. Bull.*, 1958, **55**, 316–327.

NORTON, D. W. An empirical investigation of some effects of non-normality and heterogeneity on the F-distribution. Cited by E. F. Lindquist, *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

PETERS, H. N. Multiple choice learning in the chronic schizophrenic. *J. clin. Psychol.*, 1953, **9**, 328–333.

RICKERS-OVSIANKINA, MARIA A. Studies on the personality structure of schizophrenic individuals: II. Reaction to interrupted tasks. *J. gen. Psychol.*, 1937, **16**, 179–196.

RICKSHER, C. Impressibility in dementia praecox. *Amer. J. Insan.*, 1909, **66**, 219–229.

SHAKOW, D. The nature of deterioration in schizophrenic conditions. *Nerv. ment. dis. Monogr.*, 1946, No. 70.

WEGROCKI, H. J. Generalizing ability in schizophrenia. *Arch. Psychol.*, 1940, No. 254.

WHITEMAN, M. The performance of schizophrenics on social concepts. *J. abnorm. soc. Psychol.*, 1954, **49**, 266–271.

WULFECK, W. H. Motor function in the mentally disordered: II. The relation of muscle tension to the performance of motor tasks. *Psychol. Rec.* 1941, **4**, 325–348.

(Received December 29, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 114-119

INTERPERSONAL DECISION MAKING:
THE RESOLUTION OF A DYADIC CONFLICT

JOHN M. ATTHOWE, JR.[1]
University of Alabama

INTERPERSONAL behavior is regarded as an important determinant of personality as well as of social institutions. Interaction is assumed to result in the socialization of the child (Sears, 1951), the development of both neurotic and normal behavior (Fromm, 1941; Horney, 1945; Murray, 1951; Sullivan, 1947), success or failure in psychotherapy (Fromm-Reichmann, 1949; Rogers, 1951), and emergent social systems (Parsons & Shils, 1951). If interpersonal influence can lead to both positive and negative consequences, we must be able to specify the conditions that give rise to the one or the other. Do these conditions arise within the interaction sequence, or are they logically prior to it?

Sears (1951) has suggested an operational basis for describing interpersonal influence both as an ongoing process and as an instigating condition. In the present investigation, Sear's dyadic paradigm provides the context for the study of the reciprocal interaction of two subjects faced with a decision conflict.

Conflict resolution was studied in a cooperative interpersonal setting. Two questions were posed: How do individuals arrive at joint decisions when they are faced with conflicting courses of action, and what conditions lead to more efficient or to less efficient interpersonal decisions? The time required to make a decision and its correctness, or rationality, were used as measures of efficiency.

Subjects who had previously served in an individual decision making experiment were retested in pairs. The dyadic task was identical with the individual task. Subjects were instructed to select the "better" of two alternatives; the only difference in the two experiments was the requirement that the dyadic choice be a joint (i.e., mutual) decision.

Three classes of antecedent conditions were related to the efficiency of dyadic decisions. These were: the type of choice situation or conflict, differences in the way the partners individually resolved decision conflicts, and the nature and amount of interaction.

Three types of conflict were defined; each consisted of an approach (ap) and an avoidance (av) dominant subtype. A conflict type was designated by the following ratio[2] described more fully elsewhere (Atthowe, 1960):

$$\frac{ap - av}{ap + av}$$

Ap and av represent the value of the expected outcomes, ap representing the possibility of gaining money and av that of losing. As the values of the ap and av outcomes near equality, the ratio approaches zero, a limit defined as the point of maximum risk and approximated by the ambivalent conflict. When either the ap or av outcomes dominate, the ratio nears the limit ± 1.00. Both of these limits designate minimal degrees of risk represented by the univalent conflict. A moderate degree of risk was represented by the disvalent conflict, which entails an additional feature: subjects had to reconcile the possibility of gambling (Alternative A, Table 1) with not gambling (Alternative B) in making their choice. These six situations of conflict are specified in Table 1.

The second class of antecedents concerns individual differences. The way subjects resolved similar decision conflicts on an individual basis was determined from a prior experiment (Atthowe, 1960). Subjects with different decision times and different degrees of nonrational choices were paired. The results were dyads representing different degrees of discrepancies in the way partners resolved decisions.

The third class of antecedent conditions might be better labeled emerging conditions.

[1] The author wishes to thank P. S. Siegel, P. R. Farnsworth, R. R. Sears, K. Little, and L. Festinger for reading the original draft and for their criticisms.

[2] This ratio is a means of designating types of conflict along a continuous dimension (Atthowe, 1960).

TABLE 1
TYPES OF CONFLICT
(Ap subtypes are presented to the left and av subtypes to the right)

A	B	A	B
Type I. Univalence, minimal risk			
a win 35¢ b win 26¢	win 30¢ win x¢	a lose 34¢ b lose 25¢	lose 30¢ lose x¢
Type II. Disvalence			
a win 35¢ b lose x¢	win 5¢ win 5¢	a win 26¢ b lose x¢	lose 5¢ lose 5¢
Type III. Ambivalence, maximal risk			
a win 34¢ b lose 25¢	win 26¢ lose x¢	a win 25¢ b lose 34¢	win 17¢ lose x¢

Note.—A and B are alternative choices; a and b represent possible outcomes. The probability of occurrence of any a or b was equal to one-half. x represents the unique value producing indifference of choice between A and B.

As the partners interact, differing interpersonal relationships emerge. These relations, three of which were defined prior to the experiment, were assumed to be logically independent of the measures of efficiency: (*a*) the type of influence relationship (i.e., leader-follower); (*b*) the orientation, or cohesiveness, of the dyad; and (*c*) disagreement between partners.

The influence relationship was defined by the following ratio: $\frac{X - Y}{X + Y}$. X and Y, respectively, represent the absolute difference between the individual and dyadic choices of each partner. If the joint decisions were more like the individual choices of one of the partners (e.g., X), then X was considered the leader. A ratio greater than .50 (i.e., a large change for one partner and a small change for the other) designated a strong leader-follower relationship. A ratio between .21 and .50 designated a weak relationship, and .20 or less designated a compromise relationship. Differences were calculated at the point of conflict in each of the six choice situations and summed. Conflict was defined as occurring when the alternatives were reacted to as equal in value.

The cohesiveness or orientation of the dyad was defined by the number of "we" and "I" remarks. If both partners verbalized "we" more often than "I," the dyad was designated group oriented. If they both verbalized "I" more often, the dyad was termed self-oriented. If the partners differed in their dominant remarks, the dyad was designated imbalanced (cf. Heider, 1946). The number of orientation remarks and disagreements were also recorded.

HYPOTHESES

It has been shown that different *types of conflict* yield different decision times and degrees of nonrationality in the individual choice situation (Atthowe, 1960). Would the same relations be found in the dyadic choice situation? Which of the conflict types yield more efficient and which yield less efficient dyadic decisions?

Individual differences undoubtedly play a role in determining the efficiency of joint decisions. Would the differences between partners in their typical decision speeds and their expressed nonrationality effect their joint decisions? If so, what discrepancies yield more efficient and which yield less efficient dyadic decisions?

When pairs of individuals reciprocally interact, different *interpersonal relations* are bound to emerge. Furthermore, different interpersonal relations would be expected to be associated with different joint decisions. How, then, would the efficiency of joint decisions be affected by different interpersonal relationships? What forms of interaction are found with more efficient decisions and which are found with less efficient decisions?

METHOD

The Task

The experimental task required subjects to choose the better of two alternative wagers. Each wager contained two possible monetary outcomes; the probability of each outcome was .5 and was assumed to be dependent upon the roll of an "unbiased" die. To reduce the immediate effects of reinforcement, no money was dispensed during the game. Furthermore, subjects were never told whether their choices were correct or not. The instructions emphasized the reasoning nature of the task. The wagers were said to be problems taken from the mathematical reasoning section of an advanced intelligence test and arranged as wagers. Partners were required to reach agreement before the dyadic decision could be specified.

Subjects

Fourteen pairs or dyads of male subjects were used. Each dyad was comprised of at least one subject whose

previous choices were nonrational. As a control measure, two additional dyads were tested; each was composed of two completely rational subjects.

Procedure

The wagers, printed on 3 × 5 cards, were presented as a collection of reasoning problems. The point of indifference, or conflict, was determined in a random fashion by the method of constant stimuli. All of the outcomes were held constant except one. This outcome (x in Table 1) was varied until the two wagers were seen as indifferent. The indifference point was defined as one-half the interval comprising the upper and lower limens.

The absolute values and the relative differences of the outcomes generally were held constant. The actual monetary values in the dyadic experiment differed from those in the individual setting by one to two cents. Except for the univalent conflict, the expected values of the different choice situations were equated. Consequently, the point of indifference was assumed to be "functionally comparable" for all subjects and for each of the conflict types in both the individual and dyadic settings.

Decision time and nonrationality were determined at the point of conflict. Decision time was measured in tenths of a second. Subject's individual decision time was defined as his median decision time for the six choice situations. When the mathematically expected values of the competing alternatives equaled each other, no preference could be logically indicated. This point of indifference defined a rational course of action. Any course of action corresponding to an indifference point which deviated from a rational choice was termed nonrational.

Because of the small *N* and the observed skewness, nonparametric statistics were employed. All tests of statistical significance were two-tailed.

RESULTS

Types of Conflict

Nonrationality was significantly related to the type of choice situation ($p \leq .01$). Dyadic decision times exhibited the same overall pattern, but the latter trend was not significant ($p \leq .30$) by Friedman's two-way analysis of variance. The disvalent conflict clearly instigated the least efficient dyadic decisions, and the univalent conflict exhibited maximal efficiency. The difference between the univalent and disvalent conflicts both in nonrationality and latency was significant at the .01 level.

Differences between subtypes were smaller than among types; nevertheless, av subtypes produced significantly more nonrationality than ap subtypes ($p \leq .02$). Decision times also were longer in the av situations, though not significantly so, by the Wilcoxon signed-ranks test. Conflicts characterized by the possibility of losing exhibited less efficient decisions than those characterized by the possibility of winning.

TABLE 2

DECISION MAKING EFFICIENCY AS A FUNCTION OF TYPE OF CONFLICT

(Degree is measured in cents as a deviation away from a rational solution. Time is measured in seconds.)

($N = 14$)

Type of Conflict	Nonrationality		Decision Time	
	Freq.	Degree	Avg.	Rank
Univalence				
Ap	4	8	25.7	3.7
Av	6	16	28.1	2.9
Disvalence				
Ap	10	89	38.7	2.2
Av	10	126	46.2	1.4
Ambivalence				
Ap	6	14	32.4	2.6
Av	6	16	35.1	2.3

Note.—Degree represents the total amount of nonrationality manifested. Rank is the average of the rankings for a given choice situation. A rank of 6.0 would represent the shortest decision time.

Dyadic conflict resolution differed somewhat from individual resolution. In both settings, the univalent conflict yielded the most efficiency; but, in the individual setting, ambivalence produced the largest latencies. The same overall pattern was shown in the expression of nonrationality in both settings, except individual reactions to the possibility of losing resulted in slightly more rationality. In the individual setting, decision time was the more sensitive measure of conflict. In the dyadic setting nonrationality was more sensitive.

Individual Differences

Individual differences in making decisions were found to be curvilinearly related to the efficiency of dyadic decisions. When the differences between partners were categorized into large, moderate, and small discrepancies groups, the greatest amount of nonrationality and longest decision times were found in the moderately discrepant group (see Figures 1 and 2). Larger or smaller differences between partners resulted in greater efficiency. Small differences showed slightly more efficiency than large differences. The same pattern was observed in all comparisons. When differences in individual decision times and in non-

rationality were related to both measures of dyadic efficiency, the relations approached but fell short of significance by a Kruskal-Wallis one-way analysis of variance ($p \leq .15, .25, .10, .20$, respectively).

Interpersonal Relationships

The efficiency of the dyadic decisions was also a function of the emerging relationships between partners. A weak leader-follower relationship was associated with a greater expression of nonrationality and a strong leader-follower relationship with more rationality ($p \leq .028$ by Mann-Whitney U test). Neither a strong nor a weak influence relationship was associated with the time required to reach a decision. Only one dyad exhibited compromise in their dyadic choices.

The orientation, or cohesiveness, of the dyad was not related to the efficiency of the dyadic decisions, although when the interaction sequence was divided into three phases, an early emphasis upon orientation remarks resulted in shorter decision times and a late emphasis in longer latencies ($p \leq .011$ by the Kruskal-Wallis test).

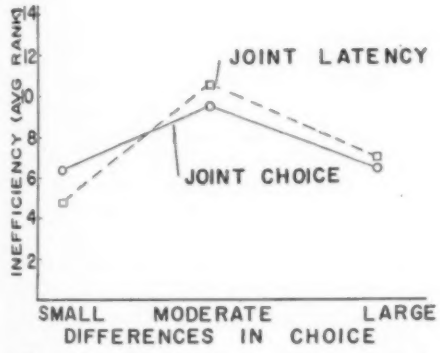

FIG. 1. The effect of differences in choice on the efficiency of decisions.

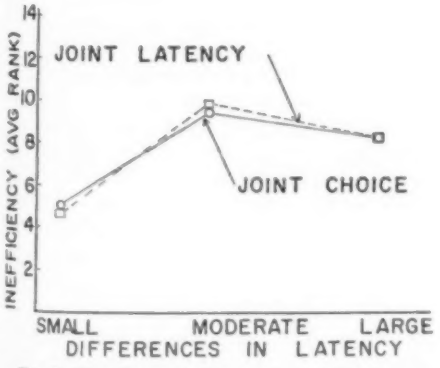

FIG. 2. The effect of differences in decision times on the efficiency of decisions.

TABLE 3

DECISION MAKING EFFICIENCY AS A FUNCTION OF TYPE OF INFLUENCE RELATIONSHIP

(Nonrationality is measured in cents and time in seconds)

Influence Relationship	N	Nonrationality		Decision Time	
		Avg. Rank	Median Value	Avg. Rank	Median Value
Strong	6	10.4	1.0	8.3	21.2
Weak	7	5.2	22.0	7.6	20.9
Compromise	1	6.0	20.0	2.0	66.7

Note.—The smaller the rank the greater the nonrationality and indecision.

The total number of orientation remarks, irrespective of their content, correlated significantly with the degree of dyadic nonrationality (rho, corrected for ties, = 0.74, $p \leq .01$). The correlation between frequency of orientation remarks and median decision times was positive but not significant (rho = 0.34). A tendency also was shown for the number of disagreements to be related to the length of the dyadic decision times (rho = 0.49, $p \leq .10$). Disagreement and nonrationality were not related.

Additional Findings

The typical strategy employed by the dyads in making decisions was one of minimizing their maximum possible loss. Dyadic decisions were significantly more conservative than extravagant ($p \leq .05$, Wilcoxon test). A nonmonotonic function, with the longest decision time coinciding with the point of conflict, was repeatedly observed as the partnerships made their decisions. As the point of indifference was approached, joint decision times lengthen ($p \leq .01$, sign test).

The two "rational" partnerships, the controls, clearly showed fewer orientation remarks (an average of 0.5 as compared to 16.5) and faster decision times (an average latency of 7.5 seconds as compared to 32.2 seconds) with much less variability.

DISCUSSION

The representativeness and generality of the results must be considered within the limits imposed by the task and instructions. A reasoning and achievement set seemed to have been established, with extensive motivational involvement as evidenced by numerous signs of sympathetic activity (e.g., flushing). One of the products of the interpersonal as compared to the individual setting was this greater involvement. The tangibleness of the numerical calculations also limited the variability of the findings, as did the cooperative nature of the task. Cooperativeness, in this sense, has been considered synonymous with cohesiveness (Deutsch, 1954). A cooperative dyad would not be expected to show the variability that a competitive one would. Considering these restrictions considerable variability and generality was shown.

The type of choice situation was an important determinant of the efficiency of dyadic decisions. Nonrationality was especially influenced by the choice situation. The possibility of choosing either a "sure thing" or a "long shot" in the disvalent situation instigated considerably more nonrationality and greater indecision. The conflict type characterized by minimal risk (the univalent situation) provided the greatest efficiency, especially the double ap-ap conflict.

The comparatively longer indecision found in the disvalent situation in the dyadic setting suggests the importance of this type of conflict for interpersonal behavior. Subjects had to reconcile their conservative and extravagant tendencies before they could reach an agreement on a choice. The instructions in the present study necessitated reciprocal action; consequently, the partners' instrumental actions were the reinforcing stimuli for further interaction. The probability that a given act of influence would yield a specifiable outcome is less certain in the interpersonal situation. When the choice situation also involves uncertainty, as in the disvalent conflict, interaction increases, resulting in longer decision times and more variability of choice.

An increase in environmental uncertainty and risk led to greater inefficiency in making decisions. These results support one of the more basic postulates of interpersonal behavior (e.g., Allport & Postman, 1945; Cantril, 1958; Killian, 1952; Sherif & Sherif, 1956). The greater inefficiency found in av as opposed to ap situations attests to the greater vigor of abient behavior.

A general similarity between individual and dyadic conflicts was also shown. The lengthening of dyadic decision time as the point of conflict (i.e., indifference) was approached points to a similar pattern of hesitancy and indecision in both settings. Furthermore, dyadic conflict resolution, like the resolution of individual conflict, proceeded conservatively. The dominant dyadic strategy was to select the alternative with the lesser possibility of losing, although this conservative trend may be specific to the university population studied (cf. Mosteller & Nogee, 1951).

If orientation remarks can be considered a representative measure of verbal interaction, interaction was inversely related to efficiency in the present study. As verbal interaction decreased, efficiency increased. Rationality was associated with infrequent interaction and stronger, or more dominating, influence. In two dyads, one of the partners was an "expert" (i.e., behaved in a rational manner); the expert not only completely influenced the choices of his partner, but he did so with a minimum of verbal interaction. Dyads containing rational partners or partnerships showed considerably less verbal interaction and disagreement as well as much greater efficiency.

Discrepancies in the way partners individually resolved conflicts were nonmonotonically related to the efficiency of their dyadic decisions. Although these findings did not reach significance, the similarity of the four relations indicates a reliable trend. Large and small discrepancies resulted in more efficient decisions. Moderate, or intermediate, differences resulted in maximal inefficiency. Too large a difference might be seen as too much to overcome; whereas very small differences could seem inconsequential. Either of these anticipations would lessen interaction. In contrast, moderate differences suggest uncertainty and risk, leading to more interaction to reduce it; consequently, decisions would take longer and result in more variability and greater nonrationality.

SUMMARY

The cooperative resolution of a dyadic conflict was studied as a special case of decision making. Two subjects were required to select a joint decision. The instructions and the task were designed to induce a reasoning set and achievement orientation. The efficiency of the joint decisions was related to the type of choice situation (conflict), the magnitude of individual differences, and the nature and amount of interpersonal influence (interaction). Latency and rationality of choice were the measures of efficiency. Maximal inefficiency was associated with the possibility of gambling, outcomes of loss rather than of gain, moderate individual differences, increasing degrees of verbal interaction, and a weak leader-follower relationship. The dyadic resolution of a decision conflict was conservative in strategy.

REFERENCES

ALLPORT, G. W., & POSTMAN, L. J. The basic psychology of rumor. *Trans. NY Acad. Sci.*, 1945, **8**, 61–81. (Reprinted in Eleanor E. Maccoby, T. M. Newcomb, & E. L. Hartley [Eds.], *Readings in social psychology*. [3rd ed.] New York: Holt, 1958. Pp. 54–65)

ATTHOWE, J. M., JR., Types of conflict and their resolution: A reinterpretation. *J. exp. Psychol.*, 1960, **59**, 1–9.

CANTRIL, H. The invasion from Mars. In Eleanor E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. (3rd ed.) New York: Holt, 1958. Pp. 291–300.

DEUTSCH, M. Field theory in social psychology. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 1. *Theory and method*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 181–222.

FROMM, E. *Escape from freedom*. New York: Farrar & Rinehardt, 1941.

FROMM-REICHMANN, FRIEDA. Recent advances in psychoanalytic therapy. In P. Mullahy (Ed.), *A study of inter-personal relations*. New York: Hermitage, 1949.

HEIDER, F. Attitudes and cognitive organization. *J. Psychol.*, 1946, **21**, 107–112.

HORNEY, KAREN. *Our inner conflicts*. New York: Norton, 1945.

KILLIAN, L. M. The significance of multiple group membership in disaster. *Amer. J. Sociol.*, 1952, **57**, 309–314.

MOSTELLER, F., & NOGEE, P. An experimental measurement of utility. *J. polit. Econ.*, 1951, **59**, 371–404.

MURRAY, H. A. Toward a classification of interactions. In T. Parsons & E. A. Shils (Eds.), *Toward a general theory of action*. Cambridge: Harvard Univer. Press, 1951. Pp. 434–464.

PARSONS, T., & SHILS, E. A. (Eds.) *Toward a general theory of action*. Cambridge: Harvard Univer. Press, 1951.

ROGERS, C. R., *Client-centered therapy*. New York: Houghton Mifflin, 1951.

SEARS, R. R. Social behavior and personality development. In T. Parsons & E. A. Shils (Eds.), *Toward a general theory of action*. Cambridge: Harvard Univer. Press, 1951. Pp. 465–478.

SHERIF, M., & SHERIF, CAROLYN W. *An outline of social psychology*. (Rev. ed.) New York: Harpers, 1956.

SULLIVAN, H. S. *Conception of modern psychiatry*. Washington, D. C.: William Alanson White Psychiatric Foundation, 1947.

(Received December 30, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 120-127

COGNITIVE CONTROLS OF ATTENTION DEPLOYMENT AS DETERMINANTS OF VISUAL ILLUSIONS[1]

RILEY W. GARDNER[2]
Menninger Foundation

THE past 10 years or so have witnessed a remarkable upsurge of interest in individual consistencies in cognitive behavior, including perception. A recent study (Gardner, Holzman, Klein, Linton, & Spence, in press) dealt with relationships between cognitive control principles—dimensions of individual consistency presumed to represent structural aspects of ego organization. Results of this study suggest that at least two of the control principles apparent in subjects' performances in laboratory and clinical tests may have particular relevance to attention and attention deployment in adaptive behavior. The first of these controls is a Field Articulation principle that is apparent in responses to situations in which subjects are confronted with perceived incongruity in the form of compelling cues relevant and irrelevant to the adaptive requirement. In the study by Gardner et al. (in press) this principle seemed to find expression in tests said to represent Field Dependence-Independence (Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954) and tests of the Constricted-Flexible Control principle (Klein, 1954). The second control is a Scanning principle that is apparent in individual consistencies in the extensiveness with which subjects characteristically sample stimuli. Both principles are relevant to situations in which the subject must attend (*a*) to stimuli outside himself or (*b*) to elements of memory schemata within himself.

The conception that such control principles represent important aspects of ego organization primarily relevant to adaptive behaviors is discussed in several publications (e.g., Gardner, 1959; Gardner et al., in press; Holzman & Klein, 1956; Klein, 1954, 1958). This view of individual consistencies in cognitive behavior is directly in line with recent emphasis in psychoanalytic theory formation upon the structures that govern adaptation (see, e.g., Hartmann, 1958; Rapaport, 1957).

In the present study, as in that of Gardner et al. (in press), attention deployment is considered both in the terms of concentration upon certain aspects of stimulus field in individual "looks" at these fields and in terms of orienting movements of the visual apparatus in the course of examining these fields. Either of these related means of controlling attention deployment may be apparent in the operation of the Field Articulation control principle.

Results of the study by Gardner et al. (in press) suggested that the dimension of individual consistencies described by Witkin (Witkin, 1959; Witkin, Karp, & Goodenough, 1959; Witkin et al., 1954) is not limited to situations requiring the subject to extract an item from a dominant field organization (as in the Embedded Figures Test). It was also apparent when the subject was required to *ignore* an item and limit response to a characteristic of the surround in a test in which the subject judged the sizes of disks bearing small pictures that tend to reduce the apparent size of the disks. (This test was originally used by Klein, 1954, to provide a measure of the Constricted-Flexible Control principle.) Subjects low in Field Articulation underestimated these disks much more than subjects high in Field Articulation. The latter could presumably concentrate attention upon the relevant disk properties and inhibit the deployment of attention to the irrelevant pictures. The capacity to deploy attention selectively was also apparent in a Free Association Test that required subject to select associations from his internal memory schemata. The performances of subjects adept at Field Articulation reflected fluent selectivity in the deployment of attention to (hypercathexis of) ideas related to the stimulus words. In contrast, protocols of subjects low in Field Articulation contained relatively few associations and frequent "blocking."

In considering the attentional behaviors

[1] The work reported here was supported by research grants from the National Institutes of Health, United States Public Health Service. Study I was supported by Grant M-1182; Study II, by Grant M-2454.

[2] The author is indebted to Gardner Murphy and John Santos for helpful readings of the manuscript.

governed by this Field Articulation principle, we are not suggesting that the entire range of individual differences observed by Witkin et al. (e.g., the generalized "passivity" and "dependence" they reported for subjects low in Field Articulation) is referable to attentional phenomena. It may well be that the individual variation in differential attention deployment apparent in laboratory and clinical tests is one facet of a much broader pattern of interrelationships that includes such general features of cognitive organization as the clarity and differentiation of self from nonself. If Piaget's (e.g., 1947) descriptions of the developmental emergence of schemata of attention deployment are adequate (above and beyond his specific hypothesis concerning attention deployment as a determinant of apparent relative magnitude), we should expect facility in the selective deployment of attention to be linked to a larger segment of personality organization.

The Scanning principle, relevant to extensiveness of sampling stimulus fields, seems to be independent of Field Articulation and to control a distinct aspect of attention deployment (see Gardner, 1959; Gardner et al., in press; Klein, 1958). The individual consistencies leading to formulation of this principle have been observed in size estimation tests and tests of incidental noticing, as well as in the "scanning" of memory schemata in the Free Association Test referred to above.

A conception of some of the processes controlled by the Scanning principle was suggested by Piaget's theory (see Piaget, 1947; Piaget, Vinh-Bang, & Matalon, 1958) of the effects of centration and decentration upon apparent relative size. Following Piaget, Gardner et al. (in press) hypothesized that extensive, balanced scanning (decentration) leads to minimal overestimation of standard disks in certain size estimation tests. Limited scanning—relative anchoring of attention upon standard (constant) stimuli in the process of comparison—was presumed to lead to larger errors. (Gardner, 1959; Gardner & Long, 1960a, 1960b; Würsten, 1947, have studied such "errors of the standard" in some detail.) The study by Gardner et al. (in press) showed that subjects characterized by unusual prominence of the defense of isolation may be among the most extensive scanners. It seemed natural that these obsessive subjects should sample widely, whether such sampling is adaptively relevant or not, and thus have small errors of the standard.

The hypothesis tested here is that the Field Articulation and Scanning principles differentially affect subjects' response to two classical types of illusion-producing figures. It was hypothesized that individual differences in Type I illusions would be predictable from core measures of Field Articulation, whereas differences in a Type II illusion would be predictable from core measures of the Scanning principle.

Type I. Tests employing the Müller-Lyer and Squares illusions were selected to represent Field Articulation because they were thought to sample individual differences in the capacity to deploy attention selectively to relevant parts of stimulus fields that also contain misleading irrelevant cues. In contrast to the Müller-Lyer, the Squares illusion is based on a "contrast" effect. These two figures thus seem to provide two different kinds of situations in which veridical judgment requires the subject to concentrate attention upon relevant lines and inhibit attention to irrelevant portions of the figures.

Benussi's (1904, 1912) ingenious studies of the effects of whole-perceiving and part-isolating attitudes upon the Müller-Lyer illusion influenced the choice of this figure for the present study. The much larger illusion effects observed by Benussi in groups of subjects instructed to attend to the entire figure, as contrasted to groups of subjects instructed to withhold attention from the angled lines, seems to provide direct evidence that attention deployment—presumably in single looks at the figure, as well as in free scanning—is a key factor determining the amount of illusion. Schiller (1942) has also distinguished between global and analytic types of response to the Müller-Lyer. In the present study it was presumed that the subjects' characteristic dispositions toward Field Articulation would produce predictable differences in response without the kind of special instruction Benussi gave to his subjects.

This view of attention deployment as a determinant of the Müller-Lyer illusion is much simpler than that offered by Piaget (see Piaget & von Albertini, 1950). He described the illusion as the final product of an elaborate

series of comparisons of lines and spaces within the figure. Both approaches, however, rest upon the assumption that attention deployment is a key variable determining the amount of illusion experienced.

It should be noted here that gestalt psychologists have spoken of the Müller-Lyer illusion in very different terms from those used by Piaget (see, e.g., Köhler & Wallach, 1944). Köhler and Fishback (1950) and Selkin and Wertheimer (1957) have employed "satiation" theory to explain the reduction of the Müller-Lyer illusion over many trials (an interpretation challenged by Eysenck & Slater, 1958). Spitz and Blackman (1958) have provided direct evidence that proneness to "satiation" effects may be determinant of the initial illusion. Therefore, a kinesthetic aftereffects test was used to provide a measure of satiation-proneness for the 80 subjects of Study I.

Type II. A second type of illusion—the Inverted *T*—was selected because of its apparent relevance to the Scanning principle. In contrast to the Müller-Lyer and Squares illusions, the Inverted *T* does not confront the subject with irrelevant lines that must be actively ignored if he is to equate parts of the figure accurately. A tendency toward extensive scanning could lead (*a*) to a small illusion effect (see Piaget et al., 1958), or (*b*) to a small error (overestimation) of the standard (also following the observations of Piaget and his coworkers). Although it was not possible to decide in advance upon the likelihood of these two possibilities, it was anticipated that the Scanning principle would have demonstrable effects upon response to this figure, and not upon response to the Müller-Lyer and Squares figures, in which the adaptive requirement is different.

METHOD

Tests Representing the Two Control Principles

Field Articulation. Embedded Figures: In Witkin's test (Witkin, 1950; Witkin et al., 1954), the subject is required to find one of eight simple figures in each of 24 complex figures. The measure used was mean log time for the 24 items. This measure was correlated .94 with the mean-time-in-seconds score used by Witkin. The corrected odd-even reliability coefficient ($N = 56$) was .95.

Rod and Frame: In this test, the subject is asked to adjust a luminous rod surrounded by a luminous frame (tilted 28° left or right) to the vertical eight times in each of three conditions: body left (28°), body right (28°), and body erect. Although the entire test was administered as it was by Witkin et al. (1954), average error for only the body-erect trials was used. Results of previous studies (e.g., Gardner et al., in press; Witkin et al., 1954) suggest that average error in the body-erect condition may be correlated somewhat more highly with related scores than is average error, total test. The corrected odd-even reliability coefficient for this measure was .92.

Scanning. Size Estimation Test A: The subject must adjust a circular light patch appearing on a square glass screen by turning a crank until the light looks equal in size to each of three hand-held 40 mm. disks: gray, 7 gms.; black, 7 gms.; gray, 67 gms. Viewing distance is 15¼ inches. The subject judges each disk four times in ADDA order from matched ascending and descending starting points. This test is similar to Size Estimation Test I used by Gardner et al. (in press). The measure for this test was subject's constant error for all three disks (12 trials). The corrected odd-even reliability coefficient was .98.

Size Estimation Test B: This test is similar to one of Klein's (1954) and to one of the size estimation procedures employed by Gardner et al. (in press), but differs in that plain rather than figured disks are presented. Using the Test A adjustment apparatus, subject judges the size of two 40 mm. disks (gray and black) presented on the wall at the left on 7 × 7-inch flesh-colored cardboard backgrounds. Viewing distance is 2 feet. The subject is allowed to turn his head freely while adjusting the variable light circle. Ascending and descending starting points are those used in Test A. The measure for this test was subject's constant error in eight judgments of the two disks. The corrected odd-even reliability coefficient was .98.

Illusions

General Procedure. Each of the illusion figures described below was presented to the subject in a room completely dark except for a roughly circular light patch approximately 25 inches in diameter that illuminated the figure being viewed. The light patch was of indistinct contour and appeared to fade gradually into the surrounding darkness. The figures consisted of strips of plastic tape, ⅛-inch wide. Except for the horizontal line of the Müller-Lyer figure, which was gray, they were covered with white tempera paint. All backgrounds were of black cardboard. Movable portions of the figures were affixed to black cardboard sliders. No subject could distinguish the contours of these sliders. The reduction of background cues undoubtedly enhanced the magnitude of the illusion effects. The subject sat in front of a table to view the figures at a distance of 10 feet. She was placed so that her vision centered at approximately the midpoint of each of the figures when accurately adjusted and was instructed not to move her head. She was allowed to move her eyes freely while making judgments. Her eyes were closed at all other times. The subject used a single control knob on the table in front of her to adjust movable portions of all four figures. This knob was connected to a selsyn apparatus on the back of each figure.

Following all experimental judgments, the subject made four control judgments of the lengths of the constant lines in the Müller-Lyer and Squares figures with the illusion-producing lines concealed. A short vertical white line served to divide the horizontal line of the Müller-Lyer figure for these judgments.

Müller-Lyer Illusion. The design of the Müller-Lyer figure was suggested by findings of several earlier experimenters (e.g., Benussi, 1904; Heymans, 1896; Judd, 1905; Szewezuk, 1938). The arrow- and featherheads were covered with white tempera paint and were angled 10° from the horizontal line, which was painted gray. In contrast to the usual Müllers-Lyer figure, the arrow- and featherheads were only 2 inches long. The light patch was just bright enough so that all subjects could see all the lines clearly. The left (constant) side of the horizontal line was 8 inches long. Turning the control knob moved the right featherhead horizontally. The subject made four judgments of the length of the left portion of the horizontal line in ADDA order from matched ascending and descending starting points. Mean adjusted sizes in the control judgments were correlated .24, $p < .05$, with the traditional mean "illusion-effect" values. The illusion effect score used consisted of the difference between the predicted size (based upon the regression of mean adjusted size, experimental judgments, on mean adjusted size, control judgments) and the mean size to which the subject actually adjusted the right-hand half of the horizontal line. The corrected odd-even reliability coefficient was .95.

Squares Illusion. This figure consisted of two pairs of squares connected by horizontal lines that extended to the centers of the inner sides. The smaller pair (1.5 inches on a side) appeared above and to the right of the larger pair (4 inches on a side). The smaller pair was connected by a 6-inch line. Turning the control knob moved the left-hand large square horizontally, changing the length of the connecting horizontal line. The subject made four judgments of the length of the horizontal line connecting the smaller squares in ADDA order from matched ascending and descending starting points. The r between control and experimental judgments was .37, $p < .01$. The score representing the illusion effect consisted of the difference between the mean size to which subject adjusted the variable line and the predicted size (based upon the regression of mean adjusted size, experimental judgments, on mean adjusted size, control judgments). The corrected odd-even reliability coefficient was .90.

Inverted T, Form A (Horizontal Standard). The standard (constant) horizontal line was 200 mm. long. The subject adjusted the length of the vertical line four times from matched ascending and descending starting points. Since no control judgments were thought possible, the traditional "illusion effect" measure for Study I consisted of the difference between the horizontal and vertical lines (e.g., 200 − 170 = 30 mm. illusion effect). The corrected odd-even reliability coefficient was .89 for the 80 subjects of Study I.

Inverted T, Form B (Vertical Standard). This figure was identical to the one described above, except that the vertical line remained constant at 200 mm., while the horizontal line was variable. Turning the control knob moved covering cardboard strips horizontally inward or outward over both ends of the horizontal line. The subject adjusted the horizontal line four times from matched ascending and descending starting points. The traditional "illusion effect" measure consisted of the difference between the mean adjusted size and 200 mm. The corrected odd-even reliability coefficient was .94 for the 68 subjects of Study II.

Gardner and Long (1960a) have shown that subjects have much greater *apparent* illusion when the vertical line is constant than when the horizontal line is the standard, because of the opposite effects of errors of the standard on the traditional "illusion effect" measures for these two forms. In Study II each subject's errors with the Inverted T figures were thus divided into two components: total error of the standard (adjusted size − 200 mm., Form B)−(200 mm. − adjusted size, Form A) and total illusion effect (adjusted size, Form B − adjusted size, Form A). Corrected odd-even reliability coefficients were: total error of the standard, .90; total illusion effect, .96.

A Test of Satiability

Kinesthetic Aftereffects Test. The test of satiation effects is modeled after those of Köhler and Dinnerstein (1947) and Wertheimer (1955). Four wooden stimulus bars are mounted on saw horses at a convenient height for a standing subject. At the front of the parallel saw horses, between which the subject stands, are a standard bar 1.5 inches wide (left) and a comparison bar with 30 $\frac{1}{16}$-inch gradations, the ninth gradation 1.5 inches wide (right). At the back of the saw horses are bars .5-inch wide (left) and 3 inches wide (right). All bars are approximately 60 inches long. The subject, blindfolded, first makes four control judgments of the 1.5-inch bar from ascending and descending starting points, in ADDA order. He then rubs the two interpolated bars (.5 left, 3.0 right) simultaneously, three times per second for 30 seconds in time with a metronome. Following this, he is required to judge the 1.5-inch bar again. After a 30-second pause, he again rubs the interpolated bars and judges the 1.5-inch test bar. This sequence is continued until four judgments of the 1.5-inch bar have been made following interpolation. The starting points for these judgments are the same as for the control judgments.

Residual scores are obviously needed to represent individual differences in satiation effects in this type of test; difference scores are not adequate. The measure used was the difference between judged size for Trial 1 following interpolation and the predicted size (based on the regression of adjusted size, Trial 1, experimental judgments, on adjusted size, Trial 1, control judgments). The r between these control and experimental judgments was .35, $p < .01$. The first trials were used because of (*a*) obvious differences in the patterning of the subjects' judgments over the four trials, and (*b*) higher test-retest reliability ($\rho = .54$) for the first trials than for the groups of four trials.

STUDY I

To evaluate the hypothesis that Field Articulation determines response to the Type

I figures, whereas Scanning controls response to the Type II figure, 80 sorority members at a small midwestern university were tested. Their ages ranged from 17 to 22. Except for 24 subjects who did not take the Embedded Figures Test, all subjects participated in the first eight procedures described above.

RESULTS

Intercorrelations among the eight measures were obtained and factor analyzed by the centroid method. Six iterations were done to improve the estimation of communalities. Four factors were extracted. The three factors accounting for most of the common variance were rotated orthogonally by the quartimax method (Neuhaus & Wrigley, 1954).[3] Table 1 contains the rotated (orthogonal) factor loadings.

Factor I

This is clearly the predicted Field Articulation factor. Like the factor obtained by Gardner et al. (in press), it seems to represent individual consistencies in the capacity to achieve differential response to relevant vs. irrelevant cues (in this case, the horizontal line vs. the illusion-producing angled lines). Low factor scores can be conceived of as representing facility in concentration in the face of compelling distractions. In the Embedded Figures Test, the subject must "find" simpler figures in contexts containing many misleading cues. In the Rod and Frame Test, he must concentrate upon the gravity cues and the rod and actively withhold attention (and response) from the frame. In the Müller-Lyer test, effective equation of the two segments of the horizontal line requires restriction of attention to the horizontal line—the same type of restriction that Benussi induced his "part-isolating" subjects to employ. In the Squares illusion, veridical judgment requires that the subject limit attention to the horizontal lines and actively avoid attending to the illusion-producing squares. Although loadings for the illusion effect measures are not high, the role of Field Articulation in solving this kind of adaptive problem seems clearly demonstrated.

[3] Rotation by the varimax method (Kaiser, 1958) produced almost exactly the same results. Steven G. Vandenberg, of the University of Michigan, provided IBM 650 programs for these solutions.

The fact that Factor I seems to combine elements of Thurstone's (1944) Factors A and B may be interpreted as follows: (*a*) Thurstone's Factor B (which included response to the Müller-Lyer figure) had a defining loading on the more difficult portion of the Gottschaldt test he used. (*b*) Factors A and B were correlated in Thurstone's study. Witkin's Embedded Figures Test, more difficult than either of Thurstone's, may be still more effective for observing Field Articulation.

Factor II

This appears to be a Scanning factor. At first glance, small errors of the standard (constant errors) in the two size estimation tests appear to be associated with large illusion effects with Form A (horizontal standard) of the Inverted *T*. Since errors of the standard and illusion effects work *against* each other in Form A, however, *apparently* large illusion effects may represent small errors of the standard. The ambiguity of this result led to Study II, which was designed to test the assumption that Factor II represents the Scanning principle, as represented in errors of the standard in all three tests.

Factor III

Although the communality of the after-effects test is rather low, this factor may represent individual differences in proneness to satiation effects (see Köhler & Wallach, 1944): the greater a subject's satiability, the greater the initial illusion with the Müller-Lyer and Squares figures. This interpretation of the loading for the Müller-Lyer judgments is in line with the findings of Spitz and Blackman (1958).

STUDY II

In order to verify the interpretation of Factor II, which rested upon the hypothesis that extensive scanners show small errors of the standard, rather than large illusion effects, in response to the Inverted *T* with horizontal standard, a new group of 68 subjects (female business college students, nursing students, and secretaries, aged 17 to 30) was tested with Size Estimation Tests A and B and *two* forms of the Inverted *T*: Form A, with horizontal standard, and Form B, with vertical standard. Half the subjects were given the two forms in

TABLE 1
ROTATED (Orthogonal) FACTOR LOADINGS: QUARTIMAX SOLUTION

Measure	Factor I	Factor II	Factor III	h^2
1. Embedded Figures: Mean Log Time	.9424	−.1006	−.0604	.9019
2. Rod and Frame: Average Error, Body Erect	.6374	−.1108	.1021	.4290
3. Size Estimation A: Constant Error	.2384	−.7180	.1470	.5940
4. Size Estimation B: Constant Error	.0708	−.8220	−.0754	.6863
5. Müller-Lyer, Illusion Effect	.4369	−.0576	−.4586	.4045
6. Squares, Illusion Effect	.3591	.0747	−.4858	.3706
7. Inverted *T*, Form A, "Illusion Effect"	.2305	.3427	.1482	.1925
8. Kinesthetic Aftereffect, Trial 1	−.0047	.0510	−.4298	.1874
Percentage of Variance Accounted for:	21.62	16.78	8.68	

each order. Starting points were changed to approximately five sigmas above and below the anticipated mean adjusted sizes, based upon adjustment of Form A by the subjects of Study I.

If the interpretation of Factor II is correct, constant error (error of the standard) in Size Estimation Tests A and B should be significantly correlated with total error of the standard, as derived from adjustments of *two* forms of the Inverted *T*. It could also be predicted that errors of the standard in the size estimation tests are negatively correlated with the traditional "illusion-effect" measure for Form A, but positively with the traditional measure for Form B (adjusted size − 200 mm.). The confirmation of these hypotheses, like the group results concerning errors of the standard reported by Gardner and Long (1960a, 1960b) would establish even more firmly the misleading nature of traditional measures for single forms of the Inverted *T* and related figures.

RESULTS

Table 2 contains intercorrelations of measures derived from the four tests. The hypothesis leading to the experiment and underlying the interpretation of Factor II as a Scanning factor is confirmed. Subjects who show large errors of the standard in the size estimation tests also show large errors of the standard in the Inverted *T*. These errors of the standard are of sufficient magnitude with the relatively large illusion figures employed to produce the predicted correlations in opposite directions between the size estimation scores and the misleading traditional "illusion effect" measures for single forms of the Inverted *T*. When illusion effect is separated from error of the standard, *r*'s between the size estimation measures and illusion effect measures are nonsignificant.

TABLE 2
PEARSON CORRELATIONS
($N = 68$)

	Inverted *T* Illusion				Size Estimation A
	Traditional "Illusion Effect" Measures		Total Error of the Standard	Total Illusion Effect	
Overestimation of Standard	Form A	Form B			Overestimation of Standard
Size Estimation A	−.20*	.24*	.36**	.07	
Size Estimation B	−.16	.20*	.30**	.05	.70***

* $p < .05$, one-tailed test.
** $p < .01$, one-tailed test.
*** $p < .001$, one-tailed test.

CONCLUSIONS

We have seen how two independent cognitive control principles—Field Articulation and Scanning—differentially guide response to two kinds of illusion figures. Although the response processes involved in the operation of these principles require further exploration, these results seem to confirm the interpretation of these controls offered by Gardner et al. (in press). Field Articulation is evident when differential response to *part* of a complex field of stimulation is required. Effective observation of Field Articulation also depends upon the *difficulty* of achieving selective attention to part of the field. For example, the difficult form of the Müller-Lyer illusion used in the present study (short, impressive angled lines close to

the horizontal line) elicits Field Articulation. Response to easier forms of the illusion may, however, be determined primarily by factors other than Field Articulation. Scanning is most evident when the adaptive situation allows expression of response dispositions concerning the extensiveness of attention deployment.

In the study by Gardner et al. (in press), a Field Articulation factor appeared most clearly for female subjects, a Scanning factor only for men. The present findings thus seem to confirm the anticipation that both principles can be demonstrated for one sex if a larger group of subjects and more effective tests of Scanning are employed.

The minimal scanner shows the kind of global and unarticulated attention deployment that characterizes children and tends to decrease with age (see Piaget, 1947, Ch. 3; Würsten, 1947). The extensive scanner apparently deploys his attention in such a way as to overcome the enhancement of magnitude Piaget assumes to be contingent upon repeated centrations on anchoring objects in the field (in this case, the standard stimuli). As Witkin's (1959) developmental data show, subjects low in Field Articulation are also similar to children, in their inability to achieve differential response to parts of stimulus fields containing conflicting information. These subjects can be conceived of as relatively inefficient in attending to relevant cues and inhibiting attention to misleading irrelevant cues. Understanding of other control principles and additional aspects of personality organization is necessary, however, to evaluation of the adequacy of an individual's cognitive controls. Extreme Field Articulation could as easily be employed in maladaptive as in adaptive ways. Extensive scanning may be an adaptive virtue in some situations and an unnecessary or confusing exertion in others. Thus the analogy between limited scanning, for example, and the attention deployment of children is not meant to imply a value judgment.

From the psychoanalytic point of view, the two experiments provide a further glimpse of two aspects of ego organization that are obviously relevant to broad classes of adaptive situations. Problems of cathexis and hypercathexis have been an important concern in psychoanalytic theory formation from the beginning. Freud (1954), for example, began to outline a view of attention deployment as early as 1895 (in the Project for a Scientific Psychology). He conceived of the shift from cathexis of the memory image of the need-satisfying object to cathexis of percepts representing external objects as a key process in the beginning differentiation of the self from the nonself. As development progresses, defense mechanisms—and we would add, cognitive controls—serve to guide attention and its deployment in enduring, individually-different ways. As Piaget has pointed out, the general emergence of schemata of attention deployment is a crucial facet of adaptive coordination of the self to the external world. The present results suggest that at least *two* distinct aspects of ego organization, in the form of cognitive controls, can be shown to represent enduring strategies of attention deployment. As in the study by Gardner et al. (in press), the apparent independence of such principles is worthy of note. In the present study, a subject adept at Field Articulation may be either a limited scanner or an extreme scanner. And these principles of cognitive control are but two of a larger number of principles that comprise the person's overall "cognitive style."

REFERENCES

BENUSSI, V. Zur Psychologie des Gestalterfassens. In A. Meinong (Ed.), *Untersuchungen zur Gegenstandstheorie und Psychologie*. Leipzig: Barth, 1904. Pp. 303–448.

BENUSSI, V. Stroboskopische Scheinbewegungen und geometrisch-optische Gestalttäuschungen. *Arch. ges. Psychol.*, 1912, **24**, 31–62.

EYSENCK, H. J., & SLATER, P. Effects of practice and rest on fluctuations in the Müller-Lyer illusion. *Brit. J. Psychol.*, 1958, **49**, 246–256.

FREUD, S. Project for a scientific psychology. In Marie Bonaparte, Anna Freud, & E. Kris (Eds.), *The origins of psychoanalysis*. New York: Basic Books, 1954. Pp. 355–445.

GARDNER, R. W. Cognitive control principles and perceptual behavior. *Bull. Menninger Clin.*, 1959, **23**, 241–248.

GARDNER, R. W., HOLZMAN, P. S., KLEIN, G. S., LINTON, HARRIET B., & SPENCE, D. P. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychol. Issues*, in press.

GARDNER, R. W., & LONG, R. I. Errors of the standard and illusion effects with the inverted-*T*. *Percept. mot. Skills*, 1960, **10**, 47–54. (a)

GARDNER, R. W., & LONG, R. I. Errors of the standard and illusion effects with *L*-shaped figures. *Percept. mot. Skills*, 1960, **10**, 107–109. (b)

HARTMANN, H. *Ego psychology and the problem of adaptation.* (Trans. by D. Rapaport) New York: International Univer. Press, 1958. (Originally published 1939)

HEYMANS, G. Quantitative Untersuchungen über das "optische Paradoxon." *Z. Psychol.*, 1896, **9**, 221–255.

HOLZMAN, P. S., & KLEIN, G. S. Motive and style in reality contact. *Bull. Menninger Clin.*, 1956, **20**, 181–191.

JUDD, C. H. Movement and consciousness. *Psychol. Monogr.*, 1905, No. **7**(1, Whole No. 29).

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187–200.

KLEIN, G. S. Need and regulation. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: Univer. Nebraska Press, 1954. Pp. 224–274.

KLEIN, G. S. Cognitive control and motivation. In G. Lindzey (Ed.), *Assessment of human motives.* New York: Rinehart, 1958. Pp. 87–118.

KÖHLER, W., & DINNERSTEIN, D. Figural after-effects in kinesthesis. In, *Miscellanea Albert Michotte.* Louvain: Institut Supérieur de Philosophie, 1947. Pp. 196–200.

KÖHLER, W., & FISHBACK, JULIA. The destruction of the Müller-Lyer illusion in repeated trials: I. An examination of two theories. *J. exp. Psychol.*, 1950, **40**, 267–281.

KÖHLER, W., & WALLACH, H. Figural aftereffects: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.

NEUHAUS, J. O., & WRIGLEY, C. F. The quartimax method: An analytic approach to orthogonal simple structure. *Brit. J. statist. Psychol.*, 1954, **7**, 81–91.

PIAGET, J. *The psychology of intelligence.* (Trans. by M. Piercy & D. E. Berlyne) London: Routledge & Kegan Paul, 1950. (Originally published 1947)

PIAGET, J., & VON ALBERTINI, BARBARA. L'illusion de Müller-Lyer. *Arch. Psychol., Genève*, 1950, **33**, No. 129, 1–48.

PIAGET, J., VINH-BANG, & MATALON, B. Note on the law of the temporal maximum of some optico-geometric illusions. *Amer. J. Psychol.*, 1958, **71**, 277–282.

RAPAPORT, D. Cognitive structures. In, *Contemporary approaches to cognition.* Cambridge: Harvard Univer. Press, 1957. Pp. 157–200.

SCHILLER, L. Ganzheitliche Auffassung und Persönlichkeitstypus. *Z. Psychol.*, 1942, **153**, 43–80.

SELKIN, J., & WERTHEIMER, M. Disappearance of the Müller-Lyer illusion under prolonged inspection. *Percept. mot. Skills*, 1957, **7**, 265–266.

SPITZ, H. H., & BLACKMAN, L. S. The Müller-Lyer illusion in retardates and normals. *Percept. mot. Skills*, 1958, **8**, 219–225.

SZEWEZUK, W. Les illusions optico-géométriques. *Trav. Lab. Psychol. Exp. Univ. Cracovie*, 1938, **2**, 89–140.

THURSTONE, L. L. *A factorial study of perception.* Chicago: Univer. Chicago Press, 1944.

WERTHEIMER, M. Figural aftereffect as a measure of metabolic efficiency. *J. Pers.*, 1955, **24**, 56–73.

WITKIN, H. A. Individual differences in ease of perception of embedded figures. *J. Pers.*, 1950, **19**, 1–15.

WITKIN, H. A. The perception of the upright. *Scient. Amer.*, 1959, **200**, 50–56.

WITKIN, H. A., KARP, S. A., & GOODENOUGH, D. R. Dependence in alcoholics. *Quart. J. Stud. Alcohol*, 1959, **20**, 493–504.

WITKIN, H. A., LEWIS, HELEN B., HERTZMAN, M., MACHOVER, KAREN, MEISSNER, PEARL B., & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

WÜRSTEN, H. L'évolution des comparaisons de longueurs de l'enfant à l'adulte. *Arch. Psychol., Genève*, 1947, **32**, 1–144.

(Received December 30, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 128-132

MODIFICATION OF BEHAVIOR THROUGH VERBAL CONDITIONING: EFFECTS IN GROUP THERAPY[1]

LEONARD P. ULLMANN, LEONARD KRASNER, AND BEVERLY J. COLLINS
Veterans Administration Hospital, Palo Alto, California

A recent conference on research in psychotherapy (Rubinstein & Parloff, 1959) pointed up the difficulties of validating the effectiveness of psychotherapy by the current methods of assessing change in behavior. The number of relevant variables in the therapy situation is such that even random assignment of cases to treatment or no treatment groups (Fairweather, Simon, Gebhard, Weingarten, Holland, Sanders, Stone, & Reahl, 1960; Rogers & Dymond, 1954) does not guarantee that differences obtained may be reasonably associated with the form of *treatment* as distinct from either the *practitioner* or the unique interaction among the variables in the situation. Rather than attempting to assess the value of the total therapy situation, it seems more efficient first to isolate the behaviors of the therapist and the patient which are characteristic of the psychotherapy situation. Then these characteristic behaviors may be systematically manipulated in order to test hypotheses about the psychotherapeutic process. Among these hypotheses are: (*a*) that one person can influence the behavior of another in a predictable direction; (*b*) that this change has a desirable effect on behavior in a second, criterion situation; and (*c*) that as far as possible, the changed behavior in the criterion situation is associated with a specific aspect of the already circumscribed behavior of the experimenter.

The first point, that one person can influence another person's verbal behavior in a specifiable manner, may be accepted as reasonably well established by the studies of operant verbal conditioning reviewed by Krasner (1958) and Salzinger (1959). It is primarily to the second point, the change of subjects' behavior in a situation other than the experimental one, that this article is addressed. Relevant studies in which behavior has been manipulated in one situation and the effect of this manipulation has been measured in a second may be roughly categorized into three types: motor operant conditioning, verbal operant conditioning, and "psychotherapy."

Motor operant conditioning techniques as a means of deliberate "therapeutic" manipulation of patient behavior has been employed in a series of recent investigations (King, Merrell, Lovinger, & Denny, 1957; Lindsley & Skinner, 1954; Peters, 1952; Peters & Jenkins, 1954; Tilton, 1957). In these studies, motor operant conditioning was purposefully used as a therapeutic technique to bring patients into more active contact with reality.

In the verbal conditioning literature, reports on generalization effects are infrequent. The usual procedure has been to present a series of test-like tasks prior to and immediately following a verbal conditioning session, and then to ascertain changes in these tests. For example, Williams (1959) used as his pre- and post-measures an "interpersonal check-list" and the Osgood Semantic Differential Scale. In one of his experimental groups, "feeling statements" were reinforced; in a second experimental group, "nonfeeling statements" were reinforced. The experimental situation itself was presented to the subjects as "psychotherapy." Although use of the reinforced verbal classes was increased during the experimental sessions, there were no significant differences between groups on the pre- and posttest measures. Using a similar design, Rogers (1960) also reported significant effects of conditioning but failure to obtain generalization as measured by a battery of four personality tests. However, Sarason (1957) did report significant generalization effects as measured by a "free word naming" situation before and after a verbal conditioning situation.

The "psychotherapy" studies generally involve the use of a series of psychological measures before and after a stated "treatment" period (Rogers & Dymond, 1954). Some studies (Grossman, 1952; Keet, 1948) make

[1] From the Psychology Research Laboratory, Veterans Administration Hospital, Palo Alto, California. This investigation was supported in part by Research Grant M-2458, National Institute of Mental Health, United States Public Health Service.

use of a specified number of experimental sessions to test the effectiveness of a given therapeutic technique. There have been several discussions (Berg, 1952; Leary & Harvey, 1956; Miller, 1954) of the rationale of the use of such tests as projective techniques, rating scales, and *Q* sorts as measures of therapeutic effectiveness. Using a different criterion and one closer to the present study, Brown (1957), working with "chronic ward patients," used as his dependent measure responsiveness to an avoidance conditioning situation (paired tone and light stimuli alternately reinforced by a painful faradic shock). After the pre-experimental measures, half the patients were subjected to various psychotherapeutic manipulations, principally group therapy. The group that had received therapy was significantly more responsive to conditioning after therapy than before, whereas the control (no therapy) group showed no significant increase after the same period of time.

Where Brown used responsiveness to conditioning as a dependent variable, the present study used verbal operant conditioning sessions as the independent variable and ratings of behavior in group therapy as the dependent variable. The purpose of this design was the attempt to test two hypotheses basic to psychotherapy: that interpersonal interaction can lead to change in a desirable manner in a second, nonexperimental situation, and that the observed change may be associated with a specific aspect of the experimenter's behavior.

METHOD

Three groups, each consisting of 10 male "continued treatment" patients at a VA neuropsychiatric hospital, served as subjects. At the start of the experiment the median age of the subjects was 35 (range: 21–47), the median education was 12 years (range: 7–16 years), and the median length of current hospitalization was 5 months (range: 1–133 months). On admission, 28 of the subjects had been diagnosed as schizophrenic reaction, one had been diagnosed as manic depressive, and one as depressive reaction. Each group was given the same experimental task, the experimenter's response being varied. All subjects were rated by group therapists before and after the experimental sessions.

Experimental task. Subjects were seen for two sessions per week for 2 weeks. The situation was structured to the patient as "a study into how people make up stories about pictures shown to them." In each session, the patients were shown four pictures and asked to make up stories to last for 5 minutes each. If responses lagged, the usual TAT instructions and encouragement were used by the experimenter. The picture stimuli used were line drawings of situations common to the experiences of patients in hospitals and were scenes such as buying a tie, mowing a lawn, fishing, buying a movie ticket, ordering a meal in a restaurant. The situations were selected to be commonplace and not emotionally arousing in themselves. This procedure minimized differential "card-pull" (Gurel & Ullmann, 1958; Ullmann, 1957a) and reduced the possibility that emotionally toned stimuli would alter verbal production in a schizophrenic sample (Ullmann, 1958). The use of less direct stimuli is in line with Krasner and Ullmann's (1958) suggestion that minimal cues (e.g., "mmm-hmm") about which the schizophrenic subject is less aware may be less threatening and therefore more effective in increasing and guiding his verbal behavior than overt examiner responses (e.g., "good-bad," "right-wrong," "true-false") which the schizophrenic may perceive as evaluative of his behavior.

Experimental groups. Each subject was assigned, in the manner described below, to one of three conditions. Once in a condition, the same examiner behavior was maintained throughout the four story-telling sessions, each of which involved four stories. That is, the pattern of examiner reinforcement started on the first trial of the first session. In the first condition, the *positive-personal* reinforcement group, the examiner responded to the verbal class to be reinforced by a head nod and the approving sound "mmm-hmm." In the second group, the *impersonal-unstructured* reinforcement group, the examiner responded by pushing a button attached to an electric counter which emitted a loud click. The dial of the counter faced the patient, and he saw an increase in the number on the dial. At the end of each story, the counter was reset to zero. Questions about the meaning of the counter were answered with the phrase that it was one of the experimental measures. In the third group, the *no reinforcement* group, the experimenter made no responses to the stories told by the subjects.

Verbal class reinforced. The verbal class reinforced was "emotional words" as defined by Ullmann and McFarland (1957). There were two reasons for the selection of this group of words. The first was the relevance of such words to group therapy ratings. Ullmann (1957b) found that ability to express feeling resulting from social premises presented in a projective test was significantly correlated with group therapy scale ratings. Similarly, Ullmann and McFarland (1957) found that the number of emotional words used in response to TAT cards was highly correlated with the group therapy ratings. In terms of relevance, then, emotional words represent the sort of material that is produced in therapy and is of particular interest to psychotherapists. An increase in the use of emotional words in the experimental situation may be therapeutic in itself, or the practice of these words during experimental sessions may lead to an increase in their use in the group therapy situation. If either or both of these possibilities were the case, it could be hypothesized that an improvement in group therapy ratings would result.

A second reason for the use of emotional words was the high rater reliability previously obtained in the

use of this class of verbal behavior. In the original work with TAT protocols (Ullmann & McFarland, 1957), rater reliability was .92. Using a single "hearing" of 16 nonreinforced 4-minute tapes of college students, three raters (including the examiner in this study) obtained a coefficient of concordance significant beyond the .001 level (Weiss, Krasner, & Ullmann, 1960).

Dependent criterion measure. The dependent measure used was the Palo Alto Group Therapy Scale (GTS) (Finney, 1954). The GTS is an 88-item scale, checked by therapists for the patient's behavior during the most recent four therapy sessions. Finney (1954) reported that for 18 groups in a neuropsychiatric hospital a median rank order correlation of .84 was obtained between scale scores and global rankings of patients by group leaders. In the same study, a rank order correlation of .80 was obtained between the average ratings by 10 ward personnel of adequacy of interpersonal relationships throughout the hospital and the patients' GTS scores.

In the present study, five group therapists[2] on two wards completed GTS forms about their patients. The therapists knew that a number of conditions were involved in the experiment but not to which one the specific patient would be assigned. One of the authors scored the ratings and then assigned patients to the three experimental groups. A restriction on absolute randomness of patient assignment was that the means and standard deviations of the preratings for the three groups were kept as equal as possible. Another of the authors (BJC) was the examiner in all cases and did not know the subjects' GTS scores. Immediately after the last experimental session, the group therapist rerated the patient.[3] In all cases, the same therapist made both the pre- and post-GTS ratings.

Results

The major results of the experiment are presented in Table 1. The means and standard deviations of the pre-experimental groups were not significantly different. The average gain ($\bar{x}_1 - \bar{x}_2$) for the positive-personal reinforcement group was 5.70 GTS points; for the impersonal-unstructured reinforcement group, it was 1.00 GTS points, and for the no reinforcement group, it was 0.50 GTS points. Finally, the own-control change was at the .05 level (one-tailed test) for the positive-personal reinforcement group ($t = 1.83$), whereas it was not significant ($t = 0.34$ and 0.15) for the other two groups.

Clinically, these results may be presented in the following manner: Five of the 10 subjects in the positive-personal reinforcement group gained 10 or more GTS points; whereas 2 of the 10 subjects in the impersonal-unstructured reinforcement group gained this much, and only 1 of the 10 subjects in the no reinforcement group gained 10 or more GTS points.

Discussion

The results of this study will be discussed in terms of hypotheses that are basic to interpreting psychotherapeutic interactions. One of these hypotheses is that the effect of such an interaction is measurable in an independent criterion situation. The frequency of use in projective situations of the verbal class conditioned in this study, emotional words, had previously been demonstrated to be significantly correlated with GTS scores (Ullmann & McFarland, 1957). The relevance of the use of emotional words to group therapy ratings may explain why positive results were obtained in this study. Further, the evidence for the validity of the criterion used in this study (Finney, 1954) and the fact that the ratings depended on actual interpersonal behavior add to the pertinence of the criterion as a measure of the effectiveness of interpersonal interactions.

In terms of amount of change, it is interesting to note the relatively brief period of time (80 minutes) involved in the experiment itself. In addition, these results were obtained within an ongoing hospital treatment program which may well have increased within-group variance.[4] Despite this evidence for the therapeutic effectiveness of verbal conditioning, however, there are considerations that limit the generalizations to be drawn. On the one

[2] The authors wish to express their thanks to George Fairweather, Ben Finney, Ralph Forsyth, Sidney Gelfand, and Donald Lim for making the criterion ratings.

[3] The time between pre- and postratings averaged 14.80 days with a standard deviation of 3.31 days.

[4] An idea of the uncontrolled variation present within an ongoing treatment program may be obtained by considering why subjects on whom pre-experimental ratings had been made did not complete the experiment: Eight of the men either failed to come to appointments, or when they came, were not cooperative in story-telling; four moved to different wards; three left the hospital; two men underwent extensive and painful dental work that interfered with story-telling; two were lost to the study when the group therapist moved to a different ward, and procedures were not completed with one man because the examiner broke her leg. In terms of the criterion pre-experimental ratings, these 20 men were not significantly different from the 30 who did complete the experiment. The average GTS rating of the 20 men who did not complete the experiment was 61.4 with a standard deviation of 20.9.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF GROUP THERAPY RATINGS BEFORE AND AFTER STORY-TELLING UNDER THREE DIFFERENT REINFORCEMENT CONDITIONS

Type of Reinforcement	N	Pre-experimental		Post-experimental		r_{12}	$\bar{x}_1 - \bar{x}_2$	t
		$\bar{x}_1$	σ_1	$\bar{x}_2$	σ_2			
positive-personal	10	58.2	20.6	63.9	25.5	.94	5.70	1.83
impersonal-unstructured	10	56.8	18.4	57.8	21.1	.91	1.00	0.43
no reinforcement	10	61.0	20.0	61.5	21.1	.89	0.50	0.15

hand, the technique is not pertinent to regressed patients who cannot tell stories. Such patients are ones who might have been successfully approached by motor operant conditioning techniques (King et al., 1957; Lindsley & Skinner, 1954). With highly verbal or more "normal" subjects—college students, for example—it is possible to hypothesize that the practice of speech in general and emotional words in particular would be less effective because these subjects have less need for additional practice in emotional expressiveness. It is therefore hypothesized that the effectiveness of verbal conditioning as a "therapeutic" technique may well be useful only with people in a particular segment of the continuum of mental health.

The results of this study also lend support to the proposition that the changes in criterion measures ascribed to psychotherapy may be associated with specifiable therapist behaviors. Evidence for this concept stems from the three experimental groups used. In both the no reinforcement and the impersonal-unstructured reinforcement groups, the amount of additional personal contact with the patients matched that of the positive-personal reinforcement group. The use of the clicker-counter in the impersonal-unstructured reinforcement group matched the additional attention called to emotional words. Such additional cuing, by itself, did not yield a significant gain in group therapy scale ratings. Therefore, within the limits of the present design, there is evidence that positive-personal reinforcing behavior of the examiner was an important factor associated with the significant change in group therapy scores.

It is possible to hypothesize that the amount and direction of change measured by the criterion should be positively correlated with change in the experimental situation in the use of the reinforced verbal class. The design of the present study did not permit a test of this hypothesis. All experimental trials were reinforced, that is, there were no nonreinforced operant trials. This was done to avoid the extinction effect on schizophrenics of unreinforced trials which might have decreased the chance of obtaining change on the dependent measure. Other reasons for the decision made in this study were the failure of Rogers (1960) and Williams (1959) to obtain generalization effects after having obtained significant conditioning, and the attempt to design an analogue of psychotherapy in which the therapist is likely to give immediate reinforcement to the patient's first emission of the desired verbal behavior. Finally, the hypothesis of the positive correlation between change in use of the verbal class during experimental sessions and change on the criterion may be oversimplified. It may well be that the important thing learned was the permissibility of spontaneously using emotional words in an interpersonal situation rather than practice in the use of the words per se.

Beyond the evidence for specific hypotheses which underlie psychotherapy, the present paper illustrates the use of verbal conditioning as a miniature situation for the study of interpersonal behavior (Krasner, 1955). Verbal conditioning may be used both as a dependent variable to measure the effects of various experimental conditions on responsiveness to reinforcement, or, as in this study, as the independent variable. In either case, the verbal conditioning situation permits the reduction and isolation of interpersonal variables with a resultant increase in the opportunity for their systematic study.

SUMMARY

This study used a verbal conditioning situation to investigate hypotheses relevant to psychotherapeutic interactions. Neuropsychiatric patients who were receiving group therapy participated in four story-telling

sessions during which emotional words were reinforced in one of the following three ways: a positive-personal manner, an impersonal-unstructured manner, and not reinforced at all. Ratings made by group therapists before and after the experimental story-telling sessions indicated a significant gain in adequacy of interpersonal relationships in group therapy for the group receiving positive-personal reinforcement. There was no significant gain for the other two groups on this criterion measure. The results support the hypotheses that one person can influence another person in a positive way and that this change may be measurable by an independent criterion situation. Further, this change in the subject's behavior may be demonstrated to be associated with specific behavior on the part of the experimenter.

REFERENCES

BERG, I. Measures before and after therapy. *J. clin. Psychol.*, 1952, **8**, 46–50.

BROWN, C. C. Changes in avoidance conditioning following psychotherapeutic treatment. *J. nerv. ment. Dis.*, 1957, **125**, 487–489.

FAIRWEATHER, G. W., SIMON, R., GEBHARD, M. E., WEINGARTEN, E., HOLLAND, J. L., SANDERS, R., STONE, G. B., & REAHL, J. Relative effectiveness of psychotherapeutic programs: A multicriteria comparison of four programs for three different patient groups. *Psychol. Monogr.*, 1960, **74**(5, Whole No. 492).

FINNEY, B. C. A scale to measure interpersonal relationships in group therapy. *Group Psychother.*, 1954, **7**, 52–66.

GROSSMAN, D. An experimental investigation of a psychotherapeutic technique. *J. consult. Psychol.*, 1952, **16**, 325–331.

GUREL, L., & ULLMANN, L. P. Quantitative differences in response to TAT cards: The relationship between transcendence score and number of emotional words. *J. proj. Tech.*, 1958, **22**, 399–401.

KEET, C. D. Two verbal techniques in a miniature counseling situation. *Psychol. Monogr.*, 1948, **62** (7, Whole No. 294).

KING, G. F., MERRELL, D. W., LOVINGER, E., & DENNY, M. R. Operant motor behavior in acute schizophrenics. *J. Pers.*, 1957, **25**, 317–326.

KRASNER, L. The use of generalized reinforcers in psychotherapy research. *Psychol. Rep.*, 1955, **1**, 19–25.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

KRASNER, L., & ULLMANN, L. P. Variables in the verbal conditioning of schizophrenic subjects. *Amer. Psychologist*, 1958, **13**, 358. (Abstract)

LEARY, T., & HARVEY, JOAN S. A methodology for measuring personality changes in psychotherapy. *J. clin. Psychol.*, 1956, **12**, 123–132.

LINDSLEY, O. R., & SKINNER, B. F. A method for the experimental analysis of the behavior of psychotic patients. *Amer. Psychologist*, 1954, **9**, 419–420. (Abstract)

MILLER, J. G. Criteria and measurement of change during psychiatric treatment. *Bull. Menninger Clin.*, 1954, **18**, 130–137.

PETERS, H. N. An experimental evaluation of learning vs. therapy in schizophrenia. *Amer. Psychologist*, 1952, **7**, 354. (Abstract)

PETERS, H. N., & JENKINS, R. L. Improvement of chronic schizophrenic patients with guided problem-solving motivated by hunger. *Psychiat. Quart. Suppl.*, 1954, **28**, 84–101.

ROGERS, C. R., & DYMOND, ROSALIND F. (Eds.) *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954.

ROGERS, J. M. Operant conditioning in a quasi-therapy setting. *J. abnorm. soc. Psychol.*, 1960, **60**, 247–252.

RUBINSTEIN, E. A., & PARLOFF, M. B. (Eds.) *Research in psychotherapy*. Washington, D. C.: American Psychological Association, 1959.

SALZINGER, K. Experimental manipulation of verbal behavior: A review. *J. gen. Psychol.*, 1959, **61**, 65–94.

SARASON, BARBARA R. The effects of verbally conditioned response classes on post-conditioning tasks. *Dissertation Abstr.*, 1957, **12**, 679. (Abstract)

TILTON, J. R. The use of instrumental motor and verbal learning techniques in the treatment of chronic schizophrenics. Unpublished doctoral dissertation, Michigan State University, 1957.

ULLMANN, L. P. Productivity and the clinical use of TAT cards. *J. proj. Tech.*, 1957, **21**, 399–403. (a)

ULLMANN, L. P. Selection of neuropsychiatric patients for group psychotherapy. *J. consult. Psychol.*, 1957, **21**, 277–280. (b)

ULLMANN, L. P. Clinical correlates of facilitation and inhibition of response to emotional stimuli. *J. proj. Tech.*, 1958, **22**, 341–347.

ULLMANN, L. P., & MCFARLAND, R. L. Productivity as a variable in TAT protocols: A methodological study. *J. proj. Tech.*, 1957, **21**, 80–87.

WEISS, R. L., KRASNER, L., & ULLMANN, L. P. Responsivity to verbal conditioning as a function of emotional atmosphere and pattern of reinforcement. *Psychol. Rep.*, 1960, **6**, 415–426.

WILLIAMS, R. I. Verbal conditioning in psychotherapy. *Amer. Psychologist*, 1959, **14**, 388. (Abstract)

(Received January 4, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 133–140

THE EFFECTS OF A COMMUNICATION ON ATTITUDE CHANGE AND SCALE JUDGMENTS[1]

WALTER WEISS
Boston University

WEISS (1959) has recently demonstrated that a person's expressed attitude can be affected by a change in his estimation of the relative favorableness-unfavorableness of a set of opinion statements. The change required was postulated as a reduction in discrimination among opinion items in the neighborhood of the range of acceptance. Subjects judging the scale values of 20 opinion items in relation to a presented standard for the middle category of judgment exhibited a striking central tendency effect, with a consequent reduction in discrimination among the items in the middle region of the scale. These subjects, in comparison with other groups given explicit end standards or no standards, extended their range of acceptance toward the punitive end of the scale. Their aggregate attitude on the issue was found to be more punitive than the positions of the other groups.

In view of this finding, it is of interest to determine if a change of attitude resulting from exposure to a persuasive communication is also attended by a change in scale judgments. Patently, a sequential effect of attitude change anteceding a change in judgment can not be demonstrated. But, at the least, evidence for the presence or absence of a relationship would be suggestive for theory.

A subsidiary purpose of the research was to examine the relevance to communication effectiveness of the subjects' evaluations of the persuasive communication. In some studies, differences in evaluation were found to be related to posttest attitude (Hovland, Lumsdaine, & Sheffield, 1949; Weiss & Fine, 1955, 1956); in others, they were not (Hovland & Mandell, 1952; Weiss, 1958). But the interpretations of such data are inevitably made uncertain by the lack of precise control groups formed by the division of the control subjects into those who would have and those who would not have given favorable evaluations of the communication. An attempt was made here to provide such judgmental control groups.

[1] The research reported in this article was completed under Grant M-1351 from the United States National Institute of Mental Health.

METHOD

A total of 141 students in summer session courses at Boston University comprise the sample. The subjects were tested during one of their regular class periods. They were told that the purpose of the study was to examine people's judgments of the opinion statements of others on a particular issue.

Booklets containing the required treatment combinations were randomly arranged by consulting a table of random numbers and were handed out individually to the students. Each booklet consisted of five sections. (The term "section" is being used only for descriptive expediency; the booklets contained no such designation.) The order for the experimental subjects was: (1) graphic rating of initial position, (2) communication followed by questions asking for evaluations of the communication, (3) or (4) judgments of the scale values of a set of opinion statements on the issue of the communication, (4) or (3) questionnaire assessing attitude on the issue, (5) semantic differential. As indicated, the order of Sections 3 and 4 were counterbalanced. For the control subjects, Section 2, containing the communication and questions evaluating it, was placed at the end of the booklet; all other sections were moved forward and the counterbalancing provision was retained. Hence, these subjects read the communication and gave their evaluations of it *only after* they had responded to the attitude questionnaire and judged the scale values of the opinion statements. All subjects were admonished not to return to a page after they had turned it; this restraint was enforced by close surveillance.

The first page of each booklet contained a nine-point graphic rating scale, which each subject marked in accord with his general position on the issue of capital punishment. Scale Position 1 was designated "strongly opposed," Position 3 "moderately opposed," Position 5 "in doubt," Position 7 "moderately favorable," and Position 9 "strongly favorable." The intervening categories were left unlabeled.

The communication was a two-page article by an anonymous author who opposed strongly and vigorously the use of capital punishment for any crime. The article was followed by a number of multiple-choice questions asking for the subject's evaluation of the fairness, truth, cogency, etc. of the communication.

The attitude questionnaire consisted of the 20 items of Form A of the Peterson-Thurstone scale of attitude toward capital punishment. The subjects responded to the questionnaire by marking one of five response

alternatives for each opinion statement: strongly agree, agree, uncertain, disagree, strongly disagree.

The subjects judged the scale values of the 36 items comprising the combined Forms A and B of the Peterson-Thurstone scales[2] by selecting a rating of 1–11 for each item depending on its judged favorableness-unfavorableness towards capital punishment. Scale Value 1 was labeled "very strong approval," 6 "neither approval nor disapproval," and 11 "very strong disapproval" of capital punishment. The subjects were told that the other numbered, but unlabeled categories represented equal steps between the midpoint category and the end ones. They were also informed that not every rating need be used, for they may feel that none of the statements is a reasonable example of a particular position on the scale. It is worth noting that the designations of the ends of the judgment scale are opposite in significance to those used for the comparable ends of the self-rating scale. Also, 20 of the items scaled comprised the attitude questionnaire and, hence, were responded to in two ways by the subjects.

The semantic differential consisted of 10 pairs of evaluative attributes drawn from the factor analytic study of Osgood and Suci (1955). A 1–7 rating scale was associated with each pair of polar adjectives in the usual manner. For half of the scales, the high value attribute designated the left end of the scale; and for half, the right end.

RESULTS AND DISCUSSION

Scale Judgments

The scale values[3] of the 36 items were determined separately for each group by the procedure recommended by Attneave and Chamblis (1957). Their method takes account of individual differences in subjective rating scales or in the utilization of the ratings for making the judgments. For each group, a U-shaped relationship was found between median scale judgments and heterogeneity of ratings for an item, as given by the reciprocal of the interquartile range, a finding that conforms with other judgment data revealing higher variability of judgment for stimuli in the middle region of a scale in comparison with those distant from the central position (Attneave & Chamblis, 1957; Edwards, 1946; Volkmann, 1951). A product-moment correlation of .995 was obtained between the scale values based on the judgments of the experimental subjects and of the control subjects. A comparable correlation was obtained for the subset of 20 items comprising the attitude questionnaire.[4] An examination of the individual scale values did not reveal differences between the two groups in their judgments of items expressing approval or disapproval of capital punishment. Furthermore, the two groups were comparable in the total number of different rating categories used in the judging and in the number of items placed in any single category or group of categories. From the data at hand and on the basis of the analyses performed,[5] there are no discernible differences between the experimental and control groups in their scale judgments of the opinion stimuli. Hence, if attitudinal differences are found as a result of exposure to the communication, they cannot be attributed to nor do they contribute to correlated judgmental effects.

Attitude Change

The mean initial position for the experimental subjects, based on the self-ratings, was 5.28; and for the control subjects, it was 5.61. The difference between the groups does not approach significance ($p > .50$).[6]

Since the groups were in no way different in their judgments of the opinion statements, the scale values of the items were computed from the responses of the entire aggregate of subjects.[7] The median scale value of the items accepted (i.e., agreed or strongly agreed with) was taken as the subject's attitude position.[8]

[2] Although each form of the attitude scale contains 20 items, the conjoint set consists of 36 different statements, since the four neutral ones of each form are identical.

[3] In the presented analyses no distinction is made between subjects responding to the attitude scale first and then judging the scale positions of the items and subjects responding to these sections in the reverse order. As regards initial position, posttest attitude, scale judgments, and evaluation of the communication, there were no differences even verging on reliability among the experimental or control subjects differentiated by this procedural variation.

[4] The original Peterson-Thurstone scale values correlated −.97 with the values obtained using the Attneave-Chamblis procedure. The sign is negative because in the original scaling the higher the number, the greater was the degree of judged approval, while the reverse obtained under the current procedure.

[5] Scaling of the items by the successive intervals procedure recommended by Edwards (1952) yielded the same results. The groups were completely comparable in the widths of the determinable scale intervals and in the measured scale values of the stimuli.

[6] All p values are two-tailed.

[7] None of the attitude results is changed if the scale values of the items are based on the responses of the control subjects alone.

[8] The attitudinal results are unchanged if such other means of determining a person's attitude are

Consequently, the higher the median score, the less favorable is the respondent to capital punishment. Table 1 contains the mean attitude for each group. By t test, the difference between the groups is significant at the .02 level, with the experimental subjects exhibiting a less favorable position. The difference between the experimental and control groups was found to be significant for those whose initial positions were in the Interval 5–7 (in doubt to moderately favorable to capital punishment) on the self-rating scale (mean 2.43 vs. mean 1.79; $p < .02$), not quite significant for those who were initially even more favorable to capital punishment, Interval 8–9 (mean 1.61 vs. mean 1.31; $p < .10$), and not different for those initially opposed to capital punishment, Interval 1–4 (mean 3.42 vs. mean 3.39).[9] Hence, on the basis of these analyses, exposure to the communication did lead to attitude change in the direction of the communicator's position. As in other studies, (Carlson, 1956; Hovland, Harvey, & Sherif, 1957; Lazarsfeld, Berelson, & Gaudet, 1948), the effect of the communication was primarily exhibited by persons whose initial positions were near the midpoint of the attitude continuum. These are prople who usually feel less strongly or are in conflict about their positions on an issue and are likely to be more responsive to the influence of persuasive communications or other change-inducing experiences.

The difference between the groups in after-attitude may be due to an increase in acceptance of statements with higher scale values (i.e., less favorable to capital punishment), a decrease in acceptance of statements with lower scale values, or both. Of the 20 items of Form A, 8 expressed sentiments favorable to capital punishment, 8 were unfavorable, and 4 were neutral. These categorizations are based on the location of an item, in terms of its median judgment, to the left or right of the midpoint category or within it. The allocation of the items to these three evaluative categories accords perfectly with the comparable significance of the original Peterson-Thurstone scale values. The mean numbers of pro, neutral, and con items accepted by each group of subjects are presented in Table 2. By t tests, the experimental group was found to have accepted significantly fewer items favoring capital punishment but more items opposing it than the control group. However, the groups were comparable in the number of neutral items accepted and in the total number accepted. Hence, for the groups as a whole, the observed difference in attitude score following exposure to the communication seems to be describable as a shift of both ends of the range of acceptance toward the unfavorable end of the scale. Furthermore, there is no apparent change in the extent of the range of acceptance. This observation can be given crude quantitative support by obtaining for each subject the range of the scale values of the items he accepts. The difference between the groups in the means of these ranges does not approach significance.

TABLE 1

POSTTEST ATTITUDE SCORES

	Experimental	Control
M	2.59	2.18
SD	0.96	1.10
N	75	66

To determine if the preceding outcomes based on the responses of individuals accord with group patterns, the percentage of each group accepting each item of Form A was determined. The relevant data are presented in Table 3.[10] The groups differed significantly in their acceptance of only 5 of the 20 items (p values $< .05$). On 2 of these items, the experimental group has a significantly smaller percentage of agreement in comparison with

utilized as: applying the original Peterson-Thurstone scale values; assigning an arbitrary score of 0–4 to each opinion response, depending on which of the five alternatives was selected, and then cumulating the scores; or assigning a score of 0–2 to each response, with no distinction made between the two "agree" categories and the two "disagree" ones, and cumulating these scores. All methods of scoring were highly intercorrelated.

[9] Bartlett's test revealed a significant heterogeneity of variances among the distributions of attitude scores of these six subgroups. The 5–7 subgroups were observed to have larger variances than the others. However, there was no reliable difference in variances between the experimental and control subjects within any subdivision based on initial position.

[10] No distinction was made between the responses of "strongly agree" and "agree" in determining the percentages accepting each item. Though not included, "uncertain" responses ranged from 1% to 28%.

TABLE 2
NUMBERS OF PRO, CON, AND NEUTRAL ITEMS ACCEPTED BY EACH GROUP

	Pro	Neutral	Con
Experimental			
M	1.68	1.37	2.63
SD	1.86	0.94	2.26
N	75	75	75
Control			
M	2.33	1.11	1.88
SD	1.80	0.91	2.14
N	66	66	66
p	$<.05$	$>.10$	.05

TABLE 3
PERCENTAGE OF EACH GROUP AGREEING WITH EACH ITEM

	Item	Scale Value	Experimental	Control
Favorable	4	0.00	0.00	1.52
	18	0.22	18.67	24.24
	11	0.57	21.33	36.36
	14	0.67	26.67	33.33
	2	0.81	33.33	42.42
	9	1.52	29.33	46.97*
	16	1.66	24.00	25.76
	5	1.66	26.67	48.48*
Neutral	12	2.03	29.33	24.24
	19	2.25	77.33	71.21
	1	2.33	6.67	0.00
	6	2.44	25.33	15.15
Unfavorable	15	2.82	5.33	9.09
	8	3.01	29.33	18.18
	10	3.44	60.00	39.39*
	3	3.55	54.67	36.36*
	17	3.80	68.00	48.48*
	20	4.06	28.00	22.73
	13	4.26	30.67	28.79
	7	4.50	28.00	21.21
			($N = 75$)	($N = 66$)

* $p < .05$.

the control group; and on 3, a higher percentage of agreement. The former 2 statements express sentiments favorable to capital punishment and the latter 3 are opposed to capital punishment. Of more importance is the fact that the 3 unfavorable items occupy consecutive rank positions as determined by their scale values and the 2 favorable items are separated by only one rank. Hence, the items on which the groups are significantly different in acceptance form almost unbroken segments of the attitudinal scale. This evidence is clearly in accord with the outcomes of the preceding analyses.

Semantic Differential

An aggregate score on the semantic differential was obtained for each subject by determining the mean of his 10 ratings. The ratings of half the scales were transformed so that on all scales the higher the number, the less favorable was the evaluation of capital punishment. The median of the aggregate distribution of scores is 5.0; and 58.1% of the experimental subjects but only 40.9% of the control subjects were above the median. The *p* value of the difference is .06. Hence, these data support the results presented in Table 1 based on the attitude questionnaire.

General Relationship between Attitude and Judgment

Since attitude was affected by the persuasive communication but scale judgments were not, it is important to determine if position on the issue of capital punishment bore any relationship to judgments of the opinion items. For, if no interconnection occurred, the judgmental comparability of the experimental and control group has diminished systematic import. Hence, comparisons were made between the judgments of subgroups of subjects distinguished as before in terms of their initial self-ratings on the issue.[11] No differences that even approached reliability could be found between experimental and control subjects within any of these subgroups. Furthermore, the obtained pattern of judgmental differences among the opinion subgroups was not affected, whether only experimental or control subjects were considered. Consequently, the distinction between such subjects was not maintained in the following analyses.

In Table 4 are presented the mean number of items allocated to the neutral category (6), to the segment of the scale designating at least some approval of capital punishment (1–5) and the segment designating disapproval (7–11). By analyses of variance, the between

[11] The results are in no way affected if after-attitude is employed instead of initial position. The correlation between initial position and posttest attitude is $-.86$ for the control subjects and $-.78$ for the experimental subjects. Also, the data patterns are not changed if only the 20 statements are examined that are common to the attitude scale and the 36-item set judged by the subjects.

groups F is significant for the neutral ($p < .05$) and approval ($p < .01$) categories. The 5–7 subgroup (initially in doubt to moderately favorable) assigns significantly fewer items to the approval category than do each of the others (p at least $< .05$) and significantly more items to the neutral category than do the initially opposed (1–4) subjects ($p < .05$) but not quite significantly more than the strongly favorable (8–9) subjects ($.20 > p > .10$).

When the median judgments of the opinion items were examined, the 5–7 subgroup was observed to have assigned higher values to at least 13 of the 16 items that expressed some degree of approval of capital punishment and lower values to at least 12 of the 16 items expressing some degree of disapproval to capital punishment than did the other two subgroups. These and the preceding results suggest a central tendency effect in the judgments of the 5–7 subgroup, an outcome that accords with Weiss' (1959) supposition concerning the effects of attitude on scale judgments of opinion statements. However, despite differences in judgments, the groups ordered the items in almost precisely the same way. All rank order correlations were .99. In general, the data bespeak a relationship between position on the issue and judgments of related opinion statements. This finding is compatible with the outcomes of some previous research (Hovland & Sherif, 1952; Sherif & Hovland, 1953; Weiss, 1959).

Taken as a whole, the results clearly reveal that a persuasive communication can influence attitude on an issue without the necessity of an attendant change in scale judgments. The preceding study by Weiss (1959) indicated that if a *certain* kind of perceptual change does occur, even without exposure to a communication, a change in expressed attitude is obtained. The two studies therefore establish that perceptual change of the kind examined is at best only a sufficient condition for attitude change. Of course, if the persuasive communication produced a much greater change in attitude, an associated effect on judgment might have been observed. And other communication variables or experimental procedures or ways of measuring change may reveal a more intertwined relationship between attitude change and perceptual change. But any expectation of a general, unqualified relation of necessity between the two is impugned by the current data.

TABLE 4

NUMBERS OF OPINION STATEMENTS ALLOCATED TO CERTAIN SEGMENTS OF THE SCALE BY SUBJECTS DIFFERING IN INITIAL POSITION

Scale Segments		Initial position		
		8–9	5–7	1–4
Favorable	*M*	16.14	14.53	17.10
	SD	3.54	3.58	3.01
Neutral	*M*	3.38	4.21	2.82
	SD	2.68	3.22	2.74
Unfavorable	*M*	16.55	17.26	16.26
	SD	3.51	3.23	3.04
	N	42	47	50

Attitude Change and Judgments of the Communication

Justification for using the control subjects' evaluations of the communication rests on the following supposition: the judgments given by these subjects are the ones they would have made had they read the communication at the same time as the experimental subjects. Although this supposition cannot be confirmed directly, discrepancy between the judgments of the control and experimental subjects would invalidate this basal requirement and congruence would be consistent with it.[12] When the subjects' evaluations were examined, the groups were not found to be significantly different on any judgment. The relevant data based on all the judgment questions are presented in Table 5.[13] Hence, the method employed appears to have yielded the necessary control groups for examining the relationship between judgment and attitude change.

On only one of the five questions was initial

[12] These implications are valid if the two groups of subjects are comparable to each other in characteristics that would affect their evaluation of the communication. The use of random assignment and the relative equivalence of the groups in initial position support this assumption.

[13] It is only for presentational convenience that each set of responses is dichotomized; for the analysis of each complete response by group table yielded a nonsignificant difference.

TABLE 5

PERCENTAGE OF EACH GROUP MAKING CERTAIN JUDGMENTS OF THE COMMUNICATION AND OF THE COMMUNICATOR'S POSITION

	Experimental	Control
Mostly or Completely One-Sided	55.4	66.1
Mostly or Entirely True	66.2	75.4
Moderately or Very Strong Case	83.8	83.1
Quite or Extremely Emotional Presentation	36.5	47.7
Author Strongly Opposed to Capital Punishment	73.0	78.5
N	74	65

TABLE 6

RELATIONSHIP BETWEEN INITIAL POSITION AND JUDGMENT OF COMMUNICATOR AS STRONGLY OPPOSED TO CAPITAL PUNISHMENT

	Initial Position		
	8–9	5–7	1–4
Control			
Percentage	87.0	60.0	86.4
N	23	20	22
Experimental			
Percentage	75.0	55.6	88.9
N	20	27	27

position found to be related to judgment of the communication. A smaller proportion of those whose initial positions were within the interval "in doubt to moderately favorable" (5–7) adjudged the communicator strongly opposed to capital punishment (the extreme alternative) than did the other subjects. The relevant data are in Table 6. For the experimental group alone, the p value of χ^2 is $<.05$ and for the control group, .06. For the two groups combined, p is $<.005$. This finding is consistent with data of Hovland, Harvey, and Sherif (1957) who offer an interpretation based on the effects of assimilation and contrast in judgment.

Since a number of the cells of the frequency table on which Table 6 is based have *N*s that are quite small, initial position could not be utilized in the general analysis of the relation between this judgment and communication effectiveness. Table 7 presents the mean posttest attitude for each judgment subgroup. By analysis of variance, only the difference between the experimental and control groups is significant. By t tests, only the difference between the "other" subgroups was significant ($p < .05$). Hence, for the aggregate of subjects this judgment was unrelated to posttest attitude.

Although the interaction term was not significant, it was thought to be sufficiently interesting to warrant further examination. Since most of the subjects giving the "other" judgment had initial positions of 5–7 and since attitude change was exhibited primarily by these subjects, their attitudes and those of the corresponding ones making the more extreme judgments were computed. The data are in Table 8. Despite the small *N*s involved, analysis of variance revealed significant *F*s for the interaction term ($p < .01$) as well as the two main effects. Patently the data do not permit specifying a directional mediating relationship between judgment of the communicator's position and attitude change. Nevertheless, greater attitude change was exhibited by those subjects who did *not* perceive the communicator as strongly opposed to capital punishment than by those who did.

TABLE 7

RELATIONSHIP BETWEEN POSTTEST ATTITUDE AND JUDGMENT OF COMMUNICATOR'S POSITION

	Strongly Opposed	Other
Experimental		
M	2.49	2.78
SD	0.97	0.94
N	54	20
Control		
M	2.21	1.96
SD	1.15	0.81
N	51	14

TABLE 8

RELATIONSHIP BETWEEN JUDGMENT OF COMMUNICATOR'S POSITION AND POSTTEST ATTITUDE AMONG SUBJECTS INITIALLY UNCERTAIN OR MODERATELY FAVORABLE TO CAPITAL PUNISHMENT

	Strongly Opposed	Other
Experimental		
M	1.97	3.01
SD	0.64	0.73
N	15	12
Control		
M	1.87	1.60
SD	0.71	0.88
N	12	8

This result does not accord with the outcomes of some previous research (Ewing, 1942; Goldberg, 1954; Hovland & Pritzker, 1957; Weiss, 1958).

As regards the other evaluative questions, posttest attitude of the experimental subjects was in no instance found related to judgment; and only on the question of the truth of the communication was it found associated with the attitudes of the control subjects. To exemplify the common attitude-judgment patterns exhibited by the subjects on these questions, data are presented in Table 9 relating to the perceived truth of the communication and attitude position. By analysis of variance, neither the main effect of judgment nor the interaction term were significant;[14] of course, the experimental group was significantly different from the control group. By t test, the attitude difference between the experimental and control subjects expressing a favorable judgment had a p value of .15, while the difference between the subjects expressing a less favorable judgment yielded a p value of $<.05$.[15] Although as previously noted the second-order difference was not significant, we can still remark that the effect of the communication on those considering it relatively untruthful was not any less than on those evaluating it as truthful. These data and the general lack of relationship between judgment and attitude change cast doubt on any assumed mediating significance of differing evaluations of a communication.

Except for one instance involving a specially selected subsample of subjects, attitude change and judgment of the communication were not interrelated. Although these data merely add to the empirical potpourri resulting from the outcomes of previous research, they are based on a more precise methodology. From a general assumption that the manner of perceiving certain qualities of a communication modulates the effectiveness of persuasion, we might expect a relationship between perception and attitude change. It may be that a certain fundamental quality not yet examined will be found to have a general and systematic connection with communicative effectiveness. Or further research may continue to exhibit variegated results. But since judgments of the communication and attitude change are both response outcomes that are methodologically intertwined, we cannot expect to determine a sequential arrangement of effects. And in the absence of precisely formulated theoretical expectations, obtained or unobtained relationships are unlikely to have particular conceptual significance by themselves.

TABLE 9
RELATIONSHIP BETWEEN POSTTEST ATTITUDE AND JUDGMENT OF TRUTH OF COMMUNICATION

	Mostly or Entirely True	Other
Experimental		
M	2.61	2.30
SD	0.98	1.07
N	49	49
Control		
M	2.48	1.73
SD	0.96	1.03
N	25	16

SUMMARY

The purposes of this study were to assess the effects of a persuasive communication on attitude change and on judgments of the scale values of opinion items and to examine the relationship between attitude change and evaluations of the communication. An experimental group of subjects was exposed to a communication advocating abolition of capital punishment, after which they evaluated the communication, judged the scale values of 36 opinion statements, and revealed their own attitudes by responding to 20 opinion items. All of the opinion statements were taken from Forms A and B of the Peterson-Thurstone scale of attitude toward capital punishment. A control group of subjects provided comparable data; but they read the communication and evaluated it *only* after having responded to the attitude scale and having judged the

[14] Even when initial position is considered, judgment is unrelated to attitude change.

[15] In the absence of control subgroups and in view of the lack of a relationship between initial position and this judgment, one might have considered testing each experimental subgroup against the total control group. The result would be a p value $< .05$ for the comparison involving the experimental subjects giving the more favorable judgment and one $> .15$ for the other subgroup. This is a reversal of the results obtained when the proper control groups are available.

scale values of the 36 items. The subjects were assigned at random to the two groups.

The major results are:

1. The communication changed attitudes in the direction advocated. But the groups were alike in their judgments of the scale values of the related opinion statements. Hence, a change in scale judgments is not a necessary condition for attitude change.

2. In general, evaluations of the communication were not related to attitude change.

REFERENCES

ATTNEAVE, F., & CHAMBLIS, D. J. An improved method for deriving equal discriminability scales from ratings. *Psychol. Bull.*, 1957, **54**, 253–255.

CARLSON, E. R. Attitude change through modification of attitude structure. *J. abnorm. soc. Psychol.*, 1956, **52**, 256–261.

EDWARDS, A. L. A critique of 'neutral' items in attitude scales constructed by the method of equal appearing intervals. *Psychol. Rev.*, 1946, **53**, 159–169.

EDWARDS, A. L. The scaling of stimuli by the method of successive intervals. *J. appl. Psychol.*, 1952, **36**, 118–122.

EWING, T. N. A study of certain factors involved in changes of opinion. *J. soc. Psychol.*, 1942, **16**, 63–88.

GOLDBERG, S. C. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.*, 1954, **49**, 325–329.

HOVLAND, C. I., HARVEY, O. F., & SHERIF, M. Assimilation and contrast effects in reactions to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55**, 244–252.

HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. D. *Experiments on mass communications*. Princeton: Princeton Univer. Press, 1949.

HOVLAND, C. I., & MANDELL, W. An experimental comparison of conclusion-drawing by the communicator and by the audience. *J. abnorm. soc. Psychol.*, 1952, **47**, 581–588.

HOVLAND, C. I., & PRITZKER, H. A. Extent of opinion change as a function of amount of change advocated. *J. abnorm. soc. Psychol.*, 1957, **54**, 257–261.

HOVLAND, C. I., & SHERIF, M. Judgmental phenomena and scales of attitude measurement: Item displacement in Thurstone scales. *J. abnorm. soc. Psychol.*, 1952, **47**, 822–832.

LAZARSFELD, P. F., BERELSON, B., & GAUDET, HAZEL. *The people's choice*. New York: Columbia Univer. Press, 1948.

OSGOOD, C. E., & SUCI, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325–338.

SHERIF, M., & HOVLAND, C. I. Judgmental phenomena and scales of attitude measurement: Placement of items with individual choice of number of categories. *J. abnorm. soc. Psychol.*, 1953, **48**, 135–141.

VOLKMANN, J. Scales of judgment and their implications for social psychology. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads*. New York: Harper, 1951. Pp. 273–294.

WEISS, W. The relationship between judgments of a communicator's position and extent of opinion change. *J. abnorm. soc. Psychol.*, 1958, **56**, 380–384.

WEISS, W. The effects on opinions of a change in scale judgments. *J. abnorm. soc. Psychol.*, 1959, **58**, 329–334.

WEISS, W., & FINE, B. J. Opinion change as a function of some intrapersonal attributes of the communicates. *J. abnorm. soc. Psychol.*, 1955, **51**, 246–253.

WEISS, W., & FINE, B. J. The effect of induced aggressiveness on opinion change. *J. abnorm. soc. Psychol.*, 1956, **52**, 109–114.

(Received January 4, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 141

CRITIQUE AND NOTES

TIME ORIENTATION AND SOCIAL CLASS: A CORRECTION

JOEL E. GREENE AND ALAN H. ROBERTS
New Mexico Highlands University

THE purpose of the present note is to evaluate the findings of a commonly cited study on time orientation (Le Shan, 1952). Le Shan hypothesized that lower class children are less future-oriented than middle class children. As a test of this hypothesis he instructed subjects representing the two social classes to "Tell me a story." The time elapsed in the story was estimated. The data as reported are shown in Table 1. Le Shan stated that the chi square value for these data was highly significant ($p = .001$, 6 *df* [p. 592]). The writers were unable to duplicate this level of significance. Rather, the value of chi square obtained from his data was 9.65 which, with 6 *df*, is not significant. In addition, it should be noted that this calculation of chi square was based upon a 2 × 7 contingency table that contained nine cells having expected frequencies less than 10; of these, six had expected frequencies less than 5.

Le Shan's apparent errors notwithstanding, the usual method of correcting for small expected frequencies is to collapse adjacent categories until acceptably large expected frequencies are obtained. There is some danger that if alternative ways of collapsing are available, scrutiny of the data may result in exploiting chance differences. Le Shan's data can be used to illustrate this point; the writers tried several combinations of categories and arrived at values ranging from $p <$ .02 to not significant. Lewis and Burke (1949) emphasize that chi squares should be based upon categories determined by criteria independent of the data to be analyzed.

When such contradictions appear in the analysis of data, equivocal interpretation of findings can be avoided only by relying upon the rationale underlying the concept of time perspective and its operational definition. In the present instance, the rationale should suggest criteria for deciding which adjacent categories should be combined. Unfortunately no solution to the problem seems to be readily available.

It appears, then, that Le Shan's findings are equivocal and should not be interpreted as clearly showing a difference in time perspective between middle and lower class children. While primary responsibility for the accurate interpretation of data remains with the researcher, it is no less the responsibility of later investigators who are dependent upon the results of a particular study for theoretical or methodological support to examine it critically. Critical examination has apparently been lacking in this case.

TABLE 1
DATA FROM LE SHAN (1952, p. 592)

	Under 1 hr.	1–12 hrs.	12–24 hrs.	1 Day–7 Days	1–2 Weeks	2 Weeks–1 Yr.	1 Yr. Plus
Lower class children ($N = 74$)	32	24	4	6	4	3	1
Middle class children ($N = 43$)	12	10	4	10	4	1	2

REFERENCES

LE SHAN, L. L. Time orientation and social class. *J. abnorm. soc. Psychol.*, 1952, **47**, 589–592.

LEWIS, D., & BURKE, C. J. The use and misuse of the chi-square test. *Psychol. Bull.*, 1949, **46**, 433–489.

(Received May 18, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 142-144

NEED ACHIEVEMENT, FEAR OF FAILURE, AND SELECTIVE RECALL[1]

WALTER R. REITMAN
Carnegie Institute of Technology

This study attempts to clarify the conceptual status of the need achievement measure developed by McClelland, Atkinson, and their colleagues by means of a further investigation of its behavioral correlates. Two experiments on selective recall of meaningful material as a function of low, moderate, or high need for achievement are described, together with a discussion of their significance in relation to the need achievement construct.

METHOD

Subjects for the first experiment were 108 male volunteers from the elementary psychology course. The 68 male subjects who participated in the second experiment were drawn the following year from the same course under an experimental participation requirement instituted that year.[2] The n Achievement distributions for the two parent populations from which subjects came were divided separately into equal thirds. Each subject then was classified as low, moderate, or high in n Achievement by referring his score to the breaking points for his parent population.

Several weeks before recall measures were administered, subjects wrote thematic apperceptive stories under Neutral Conditions. These were scored for n Achievement by experienced scorers whose scores correlated at least .90 in reliability checks with an author[3] of the system.

Achievement-Orienting instructions, describing the second session as a test of abilities in which it was important to do well, preceded recall materials for all subjects in the first experiment and half of those in the second. The other subjects in the second experiment received neutral instructions, which described the session as a routine demonstration intended to familiarize students with the types of instruments psychologists use in studying behavior. Next, subjects read instructions for the recall measure embedded in phrases similar to those of the initial instructions, so as to maintain the appropriate initial set. In these instructions, subjects were told they would have one minute to read the story to be recalled and would be asked to recall it about one half-hour later. Throughout the second experiment, all instructions and materials were in written form, to minimize interaction between the experimenter and subjects.

The stories to be recalled were paired, each story ranging from 126 to 129 words. The first two stories describe successful achievement. The central figures are working on difficult problems with determination and confidence. They succeed and feel elated. The next two stories deal with attempts to achieve that end in failure. The central figures are grim, worried, hesitant. The projects involved are based on questionable ideas, or are too difficult. The attempts fail; the protagonists give up in despair or discouragement. The next two stories are unrelated to achievement. In one, a troubled young man tries to reconcile feelings for his parents with his desire to marry. In the other, two annoyed grocers are repairing a display upset by a customer's dog so they can go home to eat and rest. The first five stories are slight modifications from ones used in previous studies (deCharms, Morrison, Reitman, & McClelland, 1955; Reitman, 1954).

All subjects were given one story only. In the first experiment, six subjects each with low, moderate, and high n Achievement scores read and recalled each one of the six stories. All subjects in the second experiment were given the first of the two failure stories.

The general recall coding system is discussed elsewhere by Reitman (1954); the recall score depends both on accuracy of detail and on general thematic content. The failure story that follows is reproduced as administered, except that the detail units have been parenthesized. The thematic content score depends on recall of the task and personal goals of the hero, of his instrumental activity, and of the outcome and the hero's affective reaction to it.

> This chap is (grim), doing some (heavy) (meditating). He is a (sophomore) and has reached an (intellectual) (crisis). He (cannot make up his mind). He is (troubled), (worried).
>
> He is (trying) to (reconcile) the (philosophies) of (Descartes) and Thomas (Aquinas)—at his (tender age) of (eighteen). He has (read) several (books on philosophy) and (feels) the (weight of the world) (on his shoulders).
>
> He (wants) to (present) a (clear-cut) (synthesis) of these two (conflicting) (philosophies), to (satisfy) his own (ego) and to gain (academic) (recognition) (from his professor).
>
> He will (mess himself up) (royally). Too (inexperienced) and (uninformed), he has (tackled) (too great) a (problem). He will (give up) in (despair) and (go down) to the (Goodyear) and (drown) his (sorrows) in a (bucket) of (Piel's).

[1] Based on a paper presented September 1959 at the APA convention. I am indebted to John W. Atkinson for his aid and encouragement, and to the National Science Foundation and the Social Science Research Council for fellowships held during the course of these studies.

[2] Both of these experiments were run in conjunction with unrelated investigations of n Achievement reported elsewhere. Full details of the common motive and performance assessment conditions are omitted here to save space. See Atkinson and Reitman (1956) and Reitman (1960).

[3] John W. Atkinson.

RESULTS

Means and *SD*s of recall scores for 108 subjects participating in the first experiment are presented in Table 1 as a function of need for achievement and type of story (success, failure, or unrelated). There are no significant main effects, but the interaction is significant ($p = .025$). The subjects with moderate n Achievement scores—previously found to be associated with evidence of "fear of failure" (e.g., McClelland, Atkinson, Clark, & Lowell, 1953), go from the highest recall scores on success and unrelated stories to the lowest scores on the failure stories. Still more interesting is the contrast between this relative inability to recall stories dealing with failure to achieve under Achievement-Orientation and the superior scores comparable subjects also moderate in n Achievement make when recalling such material under the neutral conditions of the previous studies (deCharms et al., 1955).

This result had not been predicted. This fact and the existence of procedural differences between this experiment and the previous research made a second experiment desirable. The second study directly compares relative recall of achievement failure material by subjects with moderate as opposed to high and low achievement motive scores under Achievement-Oriented and under Neutral Conditions.

As Table 2 indicates, subjects with moderate n Achievement scores once again obtain both the highest failure recall scores under Neutral Conditions and the lowest failure recall scores under Achievement-Oriented Conditions, relative to other subjects. The difference is in the expected direction, and significant at the $p = .025$ point.

TABLE 1

SELECTIVE RECALL AS A FUNCTION OF LOW, MODERATE, OR HIGH NEED ACHIEVEMENT AND SUCCESS, FAILURE, OR UNRELATED STORY ($N = 12$ in each cell)

Story Content	Need Achievement					
	Low		Moderate		High	
	M	*SD*	*M*	*SD*	*M*	*SD*
Success	16.83	4.62	19.92	5.19	18.50	4.27
Failure	22.08	5.68	16.92	6.81	13.08	4.17
Unrelated	16.75	5.51	21.50	5.38	20.83	5.96

TABLE 2

SELECTIVE RECALL OF FAILURE STORY AS A FUNCTION OF NEED ACHIEVEMENT AND ACHIEVEMENT-ORIENTED VS. NEUTRAL CONDITIONS

Condition	Need Achievement						Mean Difference in recall
	Low and High			Moderate			
	N	*M*	*SD*	*N*	*M*	*SD*	
Neutral	24	17.62	6.52	10	20.80	7.19	−3.18
Ach.-Oriented	23	19.48	6.01	11	15.73	5.17	3.75
Difference Between Mean Differences							6.93
SE of Difference							3.39
t							2.05
p (one-tailed)							.025

DISCUSSION

This replicated relationship between n Achievement score and recall of failure material under Achievement-Oriented vs. Neutral Conditions is of interest in three connections.

First as Atkinson (1958, p. vii) points out, there is as yet no generally accepted construct or system of constructs to account for differences in achievement-related imagery on the TAT, or for the many kinds of motive-performance relations obtained with thematic apperceptive measures of the achievement motive. Studies of nonideational behavior, such as performance on arithmetic tasks, frequently yield roughly linear motive-performance associations, and lead to a conception of the n Achievement score as some linear function of a need to do well.

The present results do not bear at all on this empirical utility of the n Achievement measure as a linear predictor of nonideational performance. They do, however, clearly support the theoretical position taken by McClelland et al. (1953, p. 273), Moulton (1958), and others who suggested that the n Achievement index may be a complex resultant of several distinct variables, or else may yield several distinct score categories, each of which is characterized by "emergent" qualities (e.g., "fear of failure") not linearly associated with the underlying need achievement scores.

These experiments, secondly, are relevant to research on vigilance and defense phenomena. While the present design of itself does not permit a choice between primarily perceptual and primarily mnemonic interpretations of the results, it does suggest an experimental paradigm with two basic advantages. The use of printed orienting instructions and objective indices of motive and performance makes possible exact specification of four basic predictions in such studies, namely, who will distort, what, in which directions, and under which conditions. Furthermore, while the motive and recall measures yield objective scores, they also contain surplus information. Analysis of the kinds of stories written by experimentally specified

subjects, the ways in which they respond to controlled variations in orienting instructions, and the emphases and forms of organization evident in their recall protocols should provide valuable information on the defense mechanisms involved in perceptual and mnemonic distortion.

Finally, experiments on selective recall of meaningful material seem likely to provide additional information about the complex motivational structures tapped by the n Achievement index. In so doing, they promise not only to clarify the theoretical issues involved, but also to result in still more useful motivational measures.

SUMMARY

Under Achievement-Orientation, moderate need achievement scores—previously found associated with evidence of "fear of failure"—appear related to relatively high recall of success and irrelevant stories and low recall of failure stories. A second study shows moderate n Achievement, as predicted, again associated with relatively low recall of failure stories under Achievement-Orientation, but with relatively high recall of the same stories under Neutral Orientation. From these results, need for achievement appears a complex variable yielding several distinct score categories, each of which may be characterized by "emergent" qualities (e.g., "fear of failure") not linearly associated with n Achievement scores. The paper discusses the relevance of the results to the theory and use of n Achievement measures and considers the relation of these experiments to research on vigilance and defense.

REFERENCES

ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958.

ATKINSON, J. W., & REITMAN, W. R. Performance as a function of motive strength and expectancy of goal-attainment. *J. abnorm. soc. Psychol.*, 1956, **53**, 361–366.

DECHARMS, R., MORRISON, H. W., REITMAN, W. R., & MCCLELLAND, D. C. Behavioral correlates of directly and indirectly measured achievement motivation. In D. C. McClelland (Ed.), *Studies in motivation.* New York: Appleton-Century-Crofts, 1955.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MOULTON, R. W. Notes for a projective measure of fear of failure. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958.

REITMAN, W. R. Selective recall of meaningful material as a function of achievement motivation. Unpublished masters' thesis, Wesleyan University, 1954.

REITMAN, W. R. Motivational induction and the behavioral correlates of the achievement and affiliation motives. *J. abnorm. soc. Psychol.*, 1960, **60**, 8–13.

(Received September 14, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 145–147

VALUE, MAGNITUDE, AND ACCENTUATION[1]

SAMUEL E. STAYTON AND MORTON WIENER
Clark University

Given a relationship between size estimation and value, the finding still requires theoretical explanation. Tajfel (1957) and Pastore (1949) noted that despite differences in results among the various studies concerned with the problem, larger coins—in the range from penny to half dollar—are overestimated more than smaller coins. After further analysis of the experimental evidence, Tajfel concluded that in a series of stimuli "where value changes concurrently with the dimension subjected to investigation, the differences between the stimuli of the series will be perceived as larger than the objectively equivalent differences between the stimuli of a neutral series, where no association exists between value and magnitude" (p. 193).

In all of the experiments on record the relationship between size estimation and value has always implied an overestimation of the valued object. In every instance higher value has always covaried with "greater than" or "more than" in the magnitude series, because of the particular stimuli—e.g., coins and candy—that were fortuitously selected for study. If a physical attribute dimension covaries with the value of the objects in a series, and this covariance results in an accentuation of differences in the physical attribute dimension, as suggested by Tajfel, there is no a priori reason why estimation of size and value must be positively correlated, i.e., greater value need not always result in overestimation. When the value of stimuli in a series is correlated with a decrease in the magnitude of a physical attribute, accentuation of differences in the physical attribute should result in an *underestimation* of the size of the valued object, rather than overestimation. A finding of underestimation in such an instance would lend support to the theoretical model suggested, i.e., that accentuation of a physical attribute dimension underlies the relationship of size estimation and value.

The new popularity of smaller cars permitted a test of this hypothesis.

Method

Twenty-five adults in an evening college psychology class served in a group as subjects. Subjects were first given two sheets of paper. On the first sheet were the following instructions (also read aloud by the experimenter):

> Below are six lines, three marked X–Y and three unlabeled. The lines, measured by the distance from X to Y, represent the length, width, and height of a '58 Chevrolet Bel-Aire, drawn to scale. The task is to mark off on the lines below the X–Y lines the length, width, and height of the '58 Volkswagen in relative proportion to the length, width, and height of the Chevrolet. Mark off the Volkswagen dimensions by placing a short, vertical mark to the right of the vertical line already shown, the same way as the distances of the Chevrolet dimensions were marked off.

On the second sheet subjects marked off the dimensions of the Royal portable typewriter in proportion to those of the Royal standard typewriter. This task was used in order to obscure the purpose of the experiment. The sheets were then collected.

Another set of four sheets was then distributed. The first task was a modified version of Osgood's (1957) semantic differential for the Volkswagen, consisting of five 20-point scales: positive-negative, desirable-undesirable, pleasant-unpleasant, nice-awful, and good-bad. The second task was a semantic differential like the first, except this time subjects scaled the attribute "smallness of the Volkswagen."

The third task, the "desire to own scale," had the following instructions:

> Below is a list of four objects. Indicate by a check-mark on the dotted line how much you would like to own each object at the present time. If you already own one or more of the objects, then indicate your desire to own it as if you no longer possessed the object(s). There is something else that is important. Assume that the cost of the objects or their trade-in value is not important. Pretend that you are so rich that the cost of the objects does not make any difference; the only thing that matters is that some objects are more desirable to own than others.

The four objects were the Royal standard and portable typewriters, the Volkswagen, and the Chevrolet Bel-Aire. The 15-point scale was anchored by the statements on the left: "I would not like to own one at all"; and on the right by: "I would like to own one very much."

The fourth task required the subjects to write a 25-word opinion of the Volkswagen. The purpose of this last task was to try to get a sample of the handwriting from which some estimate could be derived of the way the subject used space in general. Klein (1954) has suggested that unless characteristic styles of space utilization are considered, some relationships between size judgments and value may be obscured.

Careful consideration of the problem of order effects led to the judgment that a fixed order of presentation

[1] This report is based in part on a thesis submitted by the first author in partial fulfillment of his MA at Clark University done under the direction of the second author.

of tasks would probably result in less possible confounding—particularly between the desire to own and size judgments—than using either a random or a counterbalanced order.

RESULTS

Five scores were established for each subject. The desire to own score (the value of the Volkswagen to the subject) was determined by computing the ratio of the score marked for the Volkswagen to the average score of the other three objects. The Size Estimation score was the product of the length, width, and height judgments for the Volkswagen. The Size of Handwriting score was the number of inches occupied by the first 20 words in the description of the Volkswagen. The Value of Smallness score was derived from the five evaluative scales of the semantic differential for "smallness of the Volkswagen." A semantic differential value score for the Volkswagen was computed from the five evaluative scale responses to the car.

Twenty-four of the 25 subjects underestimated the size of the Volkswagen. The results would therefore be the same whether the size estimations scores are expressed as the subjects size judgments or as a deviation from the veridical size of the Volkswagen or from the relative size of the Chevrolet. Because no assumptions could be made about the underlying distribution of the scores, nonparametric statistical analyses were performed.

Kendall rank order correlations (tau) were computed between the various relevant scores. A tau of $-.28$ ($p < .05$, two-tailed test) was found between the desire to own the Volkswagen and the Size Estimation score; also a negative, although nonsignificant, correlation was found between the semantic differential evaluative score and size estimation (tau $= -.16$, $p > .05$, two-tailed test).[2] The greater the value of the Volkswagen to the subject, the greater the tendency to underestimate the size of the car. Since the size estimation was based on a unit relative to the given size of the Chevrolet, one could hypothesize that the underestimation found could also be a function of the greater value of the Chevrolet, i.e., the greater the desire to own the Chevrolet, the smaller the size estimation of the Volkswagen and thus overestimation of the valued object, the Chevrolet. This interpretation would require a *negative* correlation between the desire to own the Chevrolet and the size estimation of the Volkswagen. On the contrary, a positive (but nonsignificant) correlation was obtained between the desire to own the Chevrolet and estimated size of the Volkswagen (tau $= +.12$, $p > .05$). There is also a negative correlation between desire to own the Volkswagen and desire to own the Chevrolet (tau $= -.45$, $p < .01$).

A positive and significant relationship (tau $= +.37$, $p < .05$) was found between size of handwriting and the size estimation of the Volkswagen. Partialing out handwriting size to control for characteristic usage of space results in a slight but nonsignificant increase (tau $= -.31$)[3] in the relationship between size estimation and desire to own a Volkswagen. A tau of $+.277$ ($\text{tau}_{.05} = .28$, two-tailed test) was found between desire to own the Volkswagen and the Value of Smallness score of the Volkswagen derived from the semantic differential. The tau between the Value of Smallness and Size was $-.26$ ($.06 < p < .07$, two-tailed test).

DISCUSSION

The results are consistent with the hypothesis that if increasing value in a series of objects is correlated with decrease in the magnitude of a physical attribute, underestimation of the size of the valued object results, a special case of the more general hypothesis that accentuation of a correlated physical attribute is the underlying construct in accounting for judgments of a value series.

There is some evidence that the object chosen for this study met the minimal requirements to

[2] The low positive, nonsignificant correlation between the desire to own score and the semantic differential score (tau $= +.16$, $p > .05$) suggests that the two scores may be measuring different types of responses to the Volkswagen and/or that one may be a more sensitive measure of value for each subject than the other. Our prior decision to use the desire to own score as the measure of value of the Volkswagen is supported by the following considerations: (*a*) Twenty-two of the 25 subjects scored on the positive half of the evaluative continua, and 5 of these checked the highest point on all five of the evaluative scales used. In contrast, 12 subjects were above and 13 subjects below the value of one on the desire to own ratio index of value; i.e., for 12 of the 25 subjects the desire to own score for the Volkswagen was higher than the average of the scores for the other three objects and for 13 subjects it was lower. (*b*) The desire to own score assessed each subject's value of the Volkswagen in relation to judged values of three other objects. This relative judging minimized each subject's scale response bias which the semantic differential score may fail to control. (*c*) Evaluation via the semantic differential does not necessarily imply value or relevance to the person himself. On the other hand, the desire to own score in itself implies personal value and relevance.

[3] The value of tau is not comparable to that of a rho or Pearson *r*. The variance accounted for by the tau cannot be estimated as it is for the Pearson product-moment correlation (see Siegel, 1956).

test the general hypothesis. The positive relationship found between value of the object (desire to own) and the value of the physical attribute (smallness), and the negative correlation between the value of smallness and the size estimation, both suggest that for subjects in general, the value of the Volkswagen is related to its smallness. However, for some persons, smallness may be the least desirable attribute, and the Volkswagen may be desired instead for economy of operation or status or some other reason. That such is the case is suggested by the tau of .277 between desire to own and the value of smallness, which is smaller than would be expected if smallness were unequivocally positive for all subjects.

The evidence is not consistent with alternative hypotheses that the amount of underestimation of Volkswagen size is a function of greater desire to own the Chevrolet, nor that the greater the desire to own a Volkswagen the more subject should know about the Volkswagen, i.e., the more veridical should be his estimate. On the contrary, the greater the desire to own the Volkswagen the greater the deviation is the size estimate from the veridical size of the Volkswagen, and there is a positive (although nonsignificant) rather than a negative relationship between desire to own a Chevrolet and estimated size of Volkswagen. The perceptual judgment measured in this experiment is mediated by memory. Previous research (Carter & Schooler, 1949) has indicated that judgments from memory appear to accentuate the relationship between value and size estimation. Further investigation would be required to determine whether similar results would be found in a judgmental task with the object present.

Although size and weight have been the most common attribute dimensions investigated, accentuation should also result in under- or overestimation of any attribute dimension (brightness, shape of nose, etc.) correlated with value. Thus, in regard to the attribute of skin color (Marks, 1940; Secord, Bevan, & Katz, 1956) Negroes perceive their own skin color as less dark (in a culture that values "whiteness"). But sun bathers proud of their sun tan would probably accentuate the darkness of their skins. American women may accentuate the thinness of their waistlines, but in cultures where plumpness is valued, accentuation in the opposite direction would be expected. The earlier finding that valued objects are overestimated would then appear to be only a limited case of accentuation as a process hypothesized to account for a wide area of social perception and judgments.

REFERENCES

CARTER, L., & SCHOOLER, E. Value, need, and other factors in perception. *Psychol. Rev.*, 1949, **56**, 200–208.

KLEIN, G. S. Need and regulation. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1954. Pp. 224–274.

MARKS, E. Skin color judgements of Negro college students. *J. abnorm. soc. Psychol.*, 1940, **22**, 3–14.

OSGOOD, C. E. *The measurement of meaning*. Chicago: Univer. Illinois Press, 1957.

PASTORE, N. Need as a determinant of perception. *J. Psychol.*, 1949, **28**, 457–475.

SECORD, P. F., BEVAN, W., & KATZ, B. The Negro stereotype and perceptual accentuation. *J. abnorm. soc. Psychol.*, 1956, **53**, 78–83.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

TAJFEL, H. Value and the perception of magnitude. *Psychol. Rev.*, 1957, **64**, 192–204.

(Received September 15, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 148-149

DOGMATISM AND LEARNING[1]

HOWARD J. EHRLICH

Columbus Psychiatric Institute and Hospital, Ohio State University

THIS is an exploratory study of one aspect of personality and its functional significance in a learning situation. Specifically its purpose is twofold: to investigate the validity of certain aspects of the construct of dogmatism, and to demonstrate the role of personality factors in a learning situation.

The construct of dogmatism has been introduced and defined by Rokeach (1954) as "(*a*) a relatively closed cognitive organization of beliefs and disbeliefs about reality, (*b*) organized around a central set of beliefs about absolute authority which, in turn, (*c*) provides a framework for patterns of intolerance and qualified tolerance toward others." The present research is concerned with the first defining characteristic of dogmatism, its closed cognitive structure—and its logical concomitant, resistance to change. Through a quasi-experimental technique an attempt was made to demonstrate the relationship of dogmatism to the degree of learning resulting from a given classroom exposure. If the postulated "closed cognitive structure" of dogmatic persons is empirically correct, then persons high in dogmatism should exhibit less learning in a classroom situation than those low in dogmatism.

Two interrelated hypotheses have therefore been tested: Dogmatism is *inversely* related to the degree of learning in a classroom situation, and the predicted relationship between dogmatism and learning is independent of academic aptitude.

METHOD

Subjects. The original sample consisted of 100 students enrolled in four introductory sociology sections who were present for both the first and second administration of the test battery, and for whom Ohio State Psychological Examination (OSPE) scores were available from the university. A third administration, by mail, reduced the sample size to 57.

Procedure. During the first week of the academic quarter subjects were presented with the Rokeach Dogmatism Scale (1956) and a 40-item true-false test of sociological knowledge. Half the test items were statements of empirical generalizations and the other half, statements of definition. The sociology test was readministered during the last week of the quarter, an approximate time lag of 10 weeks. Five months later the Dogmatism scale and the sociology test were mailed to the original subjects. Of 62 returns, 57 were complete. Split-half reliability for the Dogmatism scale is .75; and the correlation between the two administrations of the scale, separated by 5–6 months, is .73.

RESULTS

The intercorrelations of the sociology test scores appear in Table 1. The absolute changes in the test scores for the three administrations were all significantly different from each other, using *t* tests for correlated means, at $p < .001$.

TABLE 1
INTERCORRELATIONS BETWEEN DOGMATISM, OSPE, AND SOCIOLOGY TEST SCORES

	OSPE	Sociology Test		
		Time 1	Time 2	Time 3
Dogmatism	−.28*	−.30*	−.52***	−.54***
OSPE	—	.27*	.32*	.36**
Sociology Test (I_1)		—	.66***	.63***
Sociology Test (I_2)			—	.89***

* $p < .05$.
** $p < .01$.
*** $p < .0001$.

Mail Returns vs. Nonreturns

A comparison of the 57 complete returns with the 38 nonreturns indicated a uniform pattern of mean differences. Nonreturns tended to be more dogmatic and to score lower on the OSPE, though these differences were not significant at the .05 level. They performed significantly less well on the second sociology test ($t = 3.45$; $p < .01$). The bias of the return sample in the direction of eliminating those scoring higher on dogmatism and lower on the sociology test operated against the hypothesized inverse relationship between the two by constricting their respective distributions. The data thus provided a more rigorous test of the hypotheses.

[1] The materials presented here represent a reanalysis and follow-up study on the author's master's thesis submitted to the Graduate School of the Ohio State University in 1955, and supersedes the original report. I am grateful to John Evans, Robert Stuckert, and Mark Lefton for their assistance in the collection of data. Finally I should like to express my gratitude to Melvin Seeman and Milton Rokeach for their critical reading and assistance with the original report.

TABLE 2

PARTIAL CORRELATIONS BETWEEN DOGMATISM, OSPE, AND SOCIOLOGY TEST SCORES

	Partial r with Dogmatism OSPE Constant	Partial r with OSPE Dogmatism Constant
Sociology Test (l_1)	−.24	.20
Sociology Test (l_2)	−.48**	.21
Sociology Test (l_3)	−.49**	.26

** $p < .01$.

TABLE 3

SECOND ORDER PARTIAL CORRELATIONS BETWEEN DOGMATISM, OSPE, AND SOCIOLOGY TEST SCORES

	Partial r with Dogmatism OSPE and Initial Test Scores Constant	Partial r with OSPE Dogmatism and Initial Test Scores Constant
Sociology Test (l_2)	−.43**	.12
Sociology Test (l_3)	−.45**	.18

** $p < .01$.

The Relationship between Dogmatism, OSPE, and Learning

The zero-order correlations between the three variables appear in Table 1. As expected, dogmatism yields an inverse relationship, and OSPE a direct relationship, to learning in all three time periods. In Table 2, first-order partials, controlling alternately for OSPE and dogmatism, indicate, as predicted, that (*a*) dogmatism is significantly and inversely related to learning, and (*b*) can account for a greater proportion of the variance in the sociology test scores than OSPE.

An alternative procedure for exploring the relationship between dogmatism and learning, that of testing the differences on sociology test scores for high and low dogmatic subjects, divided at the median, yielded t values for each of the three administrations that were significant beyond the .005 level. Thus subjects low in dogmatism entered the sociology classroom with a higher level of learning, learned more as a result of classroom exposure, and retained this information to a significantly greater degree than the more dogmatic subjects.

Since the subject's initial level of competence in the subject matter of the course may itself be regarded as a determinant of further learning in the classroom, and since the initial scores on the sociology test manifest only a slight relationship to academic aptitude, it is necessary to control for this as well. Second-order partial correlations controlling for initial test scores and alternately for OSPE and dogmatism are presented in Table 3. While the relationship between dogmatism and learning remains relatively unaffected, the partialing out of initial test scores decreases the correlations between OSPE and learning considerably.

DISCUSSION

The basic hypotheses of this research that dogmatism would be inversely related to the degree of learning in a classroom situation, and that such a relationship would be independent of academic aptitude have been confirmed. The study materials add further evidence to the validity of Rokeach's theoretical formulations on the nature of the dogmatic mind. Further study into the generality of the dogmatism-learning relationship seems warranted.

Finally, an incidental finding of the research that the mail returns in a student population were biased in the direction of the "better" and less dogmatic subjects suggests the need for caution in the use of mail questionnaires in such a population.

REFERENCES

ROKEACH, M. The nature and meaning of dogmatism. *Psychol. Rev.*, 1954, **61**, 194–204.

ROKEACH, M. Political and religious dogmatism: An alternative to the authoritarian personality. *Psychol. Monogr.*, 1956, **70**(18, Whole No. 425).

(Received September 24, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 150–152

SOME CONTRAST EFFECTS IN SOCIAL PERCEPTION[1]

DOUGLAS S. HOLMES AND LEONARD BERKOWITZ
University of Wisconsin

STUDENTS of social perception seem increasingly to take the position that the factors affecting a person's judgments or perceptions of other people have little (that is important, at least) in common with factors affecting his perception of inanimate objects (Tagiuri & Petrullo, 1958). Nevertheless, it can be shown that principles that affect evaluations of inanimate objects also hold for some classes of social judgments. Conditions analogous to those producing contrast and assimilation in judgments of weights also bring about these effects in judgments of another's attitudes (Hovland, Harvey, & Sherif, 1957; Peak, 1958). In these studies the subjects compared the evaluated object, whether it was alive or not, with some standard. Contrast occurred, to take an example, when the psychological distance between the object and the standard was relatively great; the subjects judged the object as being further from the standard than it actually was. On the other hand, assimilation, judging the object as being nearer to the standard than it was, took place when the standard and object were relatively close.

Generalizing from these findings, we would expect to obtain contrast effects in the perception of people whenever a person is confronted sequentially with two strangers (so that prior learning does not affect his evaluations), if the stimulus people are strikingly different from each other on some judgmental dimension. The major hypothesis of this paper is that *an anchorage or standard is formed on the basis of the experience with the first person and that subsequently contrast occurs in the perception of the second stranger when there is a great perceived difference between the two.*

In testing this view the subjects in the present experiment were required to listen to an interview between a psychologist and a female college student. This psychologist, who was either very belligerent or friendly, probably served as the standard for the subjects' subsequent series of evaluations where they rated the personalities both of another, more neutral, psychologist and of a male student from brief recordings of their voices. Since the male student is decidedly different from the "belligerent psychologist" (i.e., he is not an older, aggressive, high status professional), the immediate contrast between the two is relatively great, and the student will be evaluated as a pleasant person. The contrast is not as sharp when the "friendly psychologist" serves as the standard; he and the student are more similar behaviorally. Comparing these two conditions, we expect the student preceded by the "belligerent psychologist" to be regarded as a friendlier person than the student following the "friendly psychologist."

Experiences intervening between the presentation of the intended standard and the reception of the information to be judged also can affect the magnitude of the contrast effect. If the first psychologist is followed immediately by the second, more neutral psychologist, the standard category, "psychologist" becomes less extreme (Helson, 1948), so that there is a smaller difference between this anchorage and the male student, and the contrast effect is less. In general, then, the male student should receive a more favorable evaluation when (*a*) he follows the "belligerent psychologist," and (*b*) before rather than after judgments are made of the second, neutral psychologist.

METHOD

Subjects

Approximately 200 subjects were recruited from an introductory psychology class, and were assembled in groups ranging in size from 1–24 persons, depending upon their schedules, with the entire group receiving a given experimental treatment. In order to achieve proportional cell frequencies across experimental conditions for the analyses of variance, subjects were randomly discarded until there were 36 in each of the four main conditions, for a total of 144 subjects. There were two females to every male in each condition, but the sex variable does not have any significant effects by analysis of variance.

Procedure and Experimental Design

When the subjects arrived in the experimental room they were told the purpose of the study was to discover some of the personality factors affecting an individual's "social perceptiveness." All had previously taken a brief personality inventory. The scores on this inventory, they were informed, would be related to the accuracy with which they assessed the personalities of others on the basis of tape recordings of their voices.

The experimental conditions, organized in a factorial design, were created by means of these tape recordings. The main conditions are as follows:

[1] This research was supported by Research Grant M1540 (C2) from the National Institute of Mental Health, United States Public Health Service. This study is part of a larger program of research on factors affecting aggressive behavior.

Anchorage: Belligerent or Friendly Psychologist. The first recording presented to the subjects was that of a supposed actual interview between a psychologist in the college administration and a female student applying for a loan or scholarship. The same actors took the two parts in both versions of the interview. For half of the subjects the recording portrayed the psychologist as nonsupportive and fairly belligerent, while he acted in a much more supporting and friendly manner in the recording played to the other subjects. Since this psychologist was the predominant figure in the skit, both by virtue of the length of time he spoke and the preliminary instructions that the subject concentrate on his personality, we assume that the judgmental anchorage was largely influenced by his behavior. Immediately after the interview half of the subjects in each condition evaluated the psychologist's personality on a questionnaire and all subjects described their perceptions of him on a 28-item adjective check list. No significant differences were found on any of the dependent variables between the subjects completing the questionnaire and those not responding to this form, and this distinction is ignored in the present report.

Stimulus Objects: Psychologist and Student. Following these evaluations, all subjects listened to two one-minute-long recordings in which a second male psychologist and a male college student talked about themselves. The order of presentation of these voices was systematically varied with half of the subjects hearing the psychologist first, and the others the student first.

Dependent Variables

After each presentation, subjects attempted to describe the second psychologist and the student on a 56-item adjective check list. In addition, after they had heard both voices the subjects were given a brief, 7-item questionnaire on which they simultaneously rated both the second psychologist and the student. Two of these questions assessed the favorableness of subjects' evaluations. They are: "Would you admit him into your circle of close friends?" and "Would you admit him into the social organization (lodge, fraternity, club, sorority, etc.) of most importance to yourself?" The subjects answered by placing a check mark at the appropriate point on a line 6½ inches long, with the ends labeled as "Definitely yes" and "Definitely no." The subject's evaluation score was the distance, in one-sixteenth of an inch units, away from the favorable end, summed over both items.

RESULTS

Analyses of variance of both the adjective check list and questionnaire evaluations of the interviewing psychologist indicated that the preliminary skit was highly effective in creating the desired perceptions of this first psychologist. For both measures the "belligerent psychologist" was evaluated in a significantly more unfavorable manner than the "friendly psychologist" ($p < .01$).

TABLE 1

EVALUATIONS OF THE SECOND PSYCHOLOGIST AND THE MALE STUDENT FOLLOWING THE ANCHORAGE EXPERIENCE

Anchorage	Stimulus Order	Stimulus Person	
		Student	Psychologist
A. Adjective Check List Scores			
Belligerent	P-S	16.50_{bc}[a]	18.19_{b}
Belligerent	S-P	13.56_{d}	22.47_{a}
Friendly	P-S	15.53_{cd}	19.31_{b}
Friendly	S-P	17.08_{bc}	19.11_{b}
B. Questionnaire Scores			
Belligerent	P-S	73.56_{a}	98.33_{ab}
Belligerent	S-P	66.83_{a}	114.53_{a}
Friendly	P-S	71.64_{a}	109.08_{a}
Friendly	S-P	91.75_{b}	103.36_{a}

[a] Cells containing the same subscript are *not* significantly different from each other (at the .05 level of confidence) by Duncan multiple range test. Separate analyses were conducted for the adjective check list and questionnaire scores. For both measures the higher the mean, the more unfavorable the evaluation.

By and large, as can be seen in Table 1, the presumed comparison person affected the evaluations of at least the male student on both measuring instruments. The latter stimulus person received the most favorable evaluations when the subjects were given the experience with the "belligerent psychologist" as a comparison and they were asked to judge the student before they judged the second, neutral psychologist. Similarly, he was evaluated most unfavorably when the subjects were provided with the "friendly psychologist" as the anchorage and then heard the student's voice before they heard the voice of the second psychologist. In both instances there is a contrast effect, which apparently is sharpest when the extreme cases follow in succession.

The table also shows that there is no simple generalization of evaluations from the first to the second psychologist. Instead, here too there is an indication of judgmental contrast. The neutral psychologist received the most unfavorable evaluations when the "belligerent psychologist" served as the standard, but only if the subjects had listened to the student before hearing this stimulus person. It may well be that the second psychologist was contrasted with the extremely favorably regarded student.

DISCUSSION

The results obtained lend weight to the view that the judgmental phenomena obtained in the

psychophysics laboratory often have their counterparts in the perception and evaluation of people. Effects such as those demonstrated in this experiment seem important for at least two main reasons: (*a*), their probable operation in everyday life; and (*b*), their possible operation in psychological experiments designed to isolate quite different phenomena (such as experiments on hostility). For example, in studies of hostility displacement, the subjects' experience with the frustrator may serve as the anchorage in their judgments of other people. If these others are relatively similar to the frustrator there should be a judgmental assimilation which would result in their receiving unfavorable evaluations. On the other hand, if the others are considerably different from the perceived frustrator, there should be a contrast effect in which the others are regarded more favorably than they would have been in the absence of the anchorage experience.

REFERENCES

HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.

HOVLAND, C., HARVEY, O., & SHERIF, M. Assimilation and contrast effects in reactions to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55**, 244–252.

PEAK, HELEN. Psychological structure and psychological activity. *Psychol. Rev.*, 1958, **65**, 325–347.

TAGIURI, R., & PETRULLO, L. *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958.

(Received October 5, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 153-154

A NOTE ON ANXIETY, INSTRUCTIONS, AND WORD ASSOCIATION PERFORMANCE[1]

IRWIN G. SARASON
University of Washington

In an earlier experiment (Sarason, 1959), measures of test anxiety and general anxiety were related to the subject's performance in a situation involving two consecutive group administrations of a word association test (WAT). It was then possible to compare WAT responses on the first administration with those given on the second and also to obtain commonality scores by comparisons with a normative group. In this earlier study, half of the subjects took WAT under instructions stating that it was a sensitive personality test (experimental condition); the other subjects were merely told the rules for responding (control condition). It was predicted that the Test Anxiety Scale (TAS) (Sarason, 1958) would relate, particularly in the form of an interaction with instructions, more significantly to WAT performance than would two measures of general anxiety, Bendig's short form of Taylor's Manifest Anxiety Scale (MAS) (Bendig, 1956) and the Lack of Protection Scale (LP) (Sarason, 1958). The results of the experiment did not support this expectation; in fact, there were suggestions in the commonality analyses of greater interaction between MAS and experimental instructions than for the comparable interaction with TAS.

The present experiment was designed to follow up a problem posed by these results. The instructions originally given to the subjects stressed WAT as a personality test. The TAS items, however, deal largely with individuals' reactions to achievement and intelligence tests. If the earlier experiment were to be repeated with subjects told that WAT is a sensitive measure of intelligence rather than of personality, would the prediction of a more significant interaction of instructions with TAS than with MAS and LP be supported? The present experiment was carried out to obtain an answer to this question.

METHOD

Subjects. Subjects were 80 students in introductory psychology courses at the University of Washington. Independent of the WAT situation, subjects had been given the TAS, MAS, and LP. On each of these scales, subjects were divided at the median into high and low groups. High TAS subjects had scores of five or above; low TAS subjects had scores of four or below. High MAS subjects had scores of six or above; low MAS subjects had scores of five or below. High LP subjects had scores of eight or above; low LP subjects had scores of seven or below. Since the collection of the WAT data was independent of the personality test data, it was necessary in analyses of variance to exclude certain subjects randomly from some experimental groups in order to obtain proportionality of cell frequencies. Thus, the *N*s for the various analyses were not identical.

Procedure. All subjects received standard control instructions (Sarason, 1959) in the group administration of WAT. Control subjects received only these instructions, whereas the experimental groups also were told the following:

> As you will see, the test you are about to take is a word association test. This is one of the most sensitive tests of intelligence. From your performance on this test, we will obtain an IQ measuring your level of general intellectual ability.

Responses were untimed. When the subjects completed the first WAT, they were given the second test with instructions to give the same responses as they had given on the first administration.

RESULTS

The data were analyzed in terms of 2 × 2 analyses of variance. Three measures of WAT performance were used: The reproduction discrepancy score (RDS), defined as the number of times subject's responses on the second WAT were different from the responses given on the first administration; The Commonality-1 score, the number of times subject's responses on the first WAT agreed with the most frequent response given by Russell and Jenkins' normative group (Russell & Jenkins, 1954); The Commonality-2 score, the number of times subject's responses on the first WAT were the same as words which 100 or more subjects in the normative group gave to the stimulus.

TAS. The analysis of reproduction discrepancy scores revealed a significant effect for the variable of Instructions ($F = 4.56$, $p < .05$, df 1, 70) with higher scores in the experimental than in the control group. High and low TAS groups did not differ, although the interaction between TAS and Instructions approached significance ($F = 3.11$, $.05 < p < .10$, df 1, 70) with a tendency for greater RDSs for high TAS than low TAS subjects

[1] This research was supported by grants from the Agnes H. Anderson Fund and the United States National Institute of Mental Health (M-2397). A modification of this paper was read at the American Psychological Association meetings in Cincinnati, Ohio, September 5, 1959.

TABLE 1
MEAN REPRODUCTION DISCREPANCY SCORES (RDS) AND COMMONALITY SCORES FOR GROUP DIVIDED ON BASIS OF TEST ANXIETY SCALE (TAS) AND INSTRUCTIONS

TAS	Experimental Condition	*N*	RDS	Commonality-1	Commonality-2
High	Intelligence	19	25.53	26.79	38.16
High	Control	18	14.83	39.28	54.17
Low	Intelligence	19	18.74	33.63	48.58
Low	Control	18	17.72	30.39	44.11

under the intelligence instructions and vice versa under standard instructions. The high TAS group under the intelligence instructions showed a significantly higher mean RDS than all other groups combined ($p < .01$). The relevant data are presented in Table 1.

There was only one significant effect in each of the commonality analyses. For Commonality-1, there was a significant TAS × Instructions interaction ($F = 7.04$, $p < .01$, df 1, 70). This effect was also significant for Commonality-2 ($F = 7.55$, $p < .01$, df 1, 70). In both cases the high TAS group under intelligence instructions showed lower commonality scores than the comparable low TAS groups. Under standard instructions, the tendency was for the high TAS subjects to obtain higher commonality scores.

MAS and LP. For both MAS and LP, there were higher RDSs under the intelligence instructions than under the standard condition, but neither the MAS nor the LP analyses showed any effects due to anxiety or anxiety × instructions.

DISCUSSION

This study began with the question of whether TAS interacts with instructions more significantly than MAS and LP. The answer appears to be affirmative. As in the earlier experiment, the TAS × Instructions interaction tended toward significance when the measure was the number of disagreements between subject's responses to two word association tests (RDS). However, the commonality measures of subject's agreement with frequent response words in a normative group yielded highly significant interactions in the present study but not in the earlier one.

In the earlier experiment, there were significant effects due to both MAS and MAS × Instructions for the RDS and commonality measures. No such effects were observed here. While previous research (Sarason, 1956) has shown MAS to relate significantly to certain kinds of intelligence instructions, the nature of the task and the actual instructions used may be highly important. For the present task, it would appear likely that the change in experimental instructions from a personality to an intelligence emphasis contributed to the absence of significant effects involving MAS. A similar explanation seems appropriate for the LP results.

The present TAS results support the hypothesis that the closer the nature of the experimental stress is to the content of anxiety items, the more predictive the items are. One interpretation is that under the intelligence instructions, subjects with high TAS scores tended to become threatened and disorganized and emitted relatively more interfering and idiosyncratic self-centered responses than did other subjects. This tendency was reflected in the greater number of relatively unusual or less frequently occurring responses and greater RDSs in the high TAS intelligence group than in the high TAS control group. No similar differences were found for the experimental and control low TAS groups.

The findings of both the earlier experiment and the present one are consistent with a construct of anxiety conceived as a class of self-oriented responses which interfere with ongoing performance. Anxiety scales in this view might be considered as devices for selecting individuals who differ in the tendency to emit these responses under appropriate motivational conditions.

REFERENCES

BENDIG, A. W. The development of a short form of the manifest anxiety scale. *J. consult. Psychol.*, 1956, **20**, 384.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. *USN Off. Naval Res. tech. Rep.*, 1954, No. 11. (Contract No. N8)

SARASON, I. G. Effect of anxiety, motivational instructions, and failure on serial learning. *J. exp. Psychol.*, 1956, **51**, 253–260.

SARASON, I. G. Interrelationships among individual difference variables, behavior in psychotherapy, and verbal conditioning. *J. abnorm. soc. Psychol.*, 1958, **56**, 339–344.

SARASON, I. G. Relationships of measures of anxiety and experimental instructions to word association test performance. *J. abnorm. soc. Psychol.*, 1959, **59**, 37–42.

(Received October 7, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 155–157

SOME VERBAL ASPECTS OF PRIMARY PROCESS THOUGHT IN SCHIZOPHRENIA[1]

ALVIN G. BURSTEIN[2]
University of Chicago

In connection with his study of dreams, Freud developed the concept of primary process, a mode of mental functioning possessing characteristics much different from those of typical waking thought. Primary process has been said to characterize not only dreams but also the thinking of schizophrenics. Freud (1925) specifically related primary process to the verbal behavior of schizophrenics when he stated, "In schizophrenia *words* are subject to the same process as that which makes dream images out of dream thought, the one we have called the primary mental process" (p. 131).

The present study attempts to test the hypothesis that the effects of primary process are evident in the verbal behavior of schizophrenics. Two aspects of primary process thought were chosen: the treatment of antonyms as synonyms, and the treatment of homonyms as synonyms.

METHOD

Antonyms. Freud (1949), describing one aspect of primary process thought, said, "... contraries are not kept apart from each other, but are treated as though they were identical" (p. 53). This confusion of opposites and identicals is specifically related to the area of language by Freud (1956): "... the most ancient languages behave exactly like dreams in this respect ... (strong-weak, old-young, far-near, bind-sever), ... " (p. 318).

As a measure of this presumed tendency, a task was developed consisting of 97 multiple-choice items, on each of which subject was instructed to choose the synonym for a given word from among three alternatives consisting of a synonym, an antonym, and a third alternative unrelated to the word to be defined. This third type of alternative is called here the unidentified error alternative. An example will serve to illustrate:

Circle the word that means almost the same as BIG:
Large (syn.) Small (ant.) Individual (ue.)

The purpose of the unidentified error alternative is to measure the number of antonym markings made upon bases other than those anticipated by the hypothesis, e.g., obscure communalities, idiosyncratic associations, random markings, etc. Since choices made upon such unanticipated bases would by chance result as often in marking the antonym alternative as in marking the unidentified error alternative itself, a score of antonym markings minus unidentified error markings would represent a relatively pure measure of the number of choices of antonyms per se. Another way of putting it is that there should be as many errors made for unidentified reasons hidden within the antonym markings as there are markings of the unidentified error alternatives themselves, so that subtracting out the unidentified error markings yields a corrected measure of sheer choice of antonyms per se.

In order to control for the effect upon choices of frequency of occurrence in the language, the antonyms and unidentified error alternatives were roughly equated in the task as a whole by Thorndike and Lorge's (1944) summary of four different word frequency counts.

To control for the effects of position markings, each type of alternative appeared equally often in each of the three possible positions, with the possible sequences of alternatives appearing randomly.

Homonyms. Freud gave many examples of primary process in which similarity of sound is treated as similarity of meaning. For example, a dreamer is said to have substituted "ears" (Ahren) of barley and wheat for "honor" (Ehren) on the basis of their being homonyms (1956, p. 406).

Accordingly, a task was constructed consisting of 48 items parallel in form to those discussed above. Each item required the subject to choose the *synonym* for a given word from among three alternatives, a synonym, a homonym, and an unidentified error alternative. The following sample is illustrative:

Circle the word that means almost the same as SEA
Ocean (syn.) See (hom.) Idol (ue.)

This design presupposes, of course, that in reading the items silently the subject is aware of how the words would sound if read aloud. If this were not the case, the subject could not respond on the basis of two words as homonyms.

Once again, the purpose of the unidentified error alternative was to measure choices of the predicted error that were made upon bases other than those anticipated. In a like manner, an error score of homonym markings minus unidentified error markings would yield a relatively pure measure of the treatment of homonyms as synonyms. In order to control for the effects of frequency of occurrence in the language, each

[1] This paper is based upon a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, University of Chicago, 1959. The study was supported in part by Research Grant M-1416 from the National Institute of Mental Health, United States Public Health Service.

[2] The author is indebted to the members of his graduate study committee, and especially to his sponsor, Loren J. Chapman, for invaluable guidance at every stage of the research.

Currently on the staff of the Department of Psychology at the University of Michigan.

homonym appeared also as an unidentified error alternative in another item.

We may now state the hypothesis in operational terms: Schizophrenic subjects when instructed to choose synonyms, exceed normal subjects in their choice of homonyms and antonyms as measured by the difference score of predicted error markings minus unidentified error markings.

Subjects. The tasks described above were administered to 26 chronic schizophrenics and to 24 normal control subjects. The CVS, a brief verbal intelligence test, (Hunt, French, Klebanoff, Mensh, & Williams, 1948) was given to each subject. All subjects were Caucasians whose native tongue was English, and who gave no evidence in either their records or in interviews of central nervous system pathology, drug addiction, alcoholism, or mental deficiency.

The schizophrenics had been hospitalized continuously for at least 5 years with a diagnosis of schizophrenia or dementia praecox and had not received chemotherapy within the previous month or shock therapy within the previous year. Of the normal controls, 7 were recruited from ambulatory medical and surgical patients at a general hospital, 12 were employees of a large Chicago hotel, and the remaining 5 were Chicago firemen.

In all cases the CVS was administered either individually or in groups of two or three. The subjects were required to answer all questions on the paper-and-pencil tasks. Both the schizophrenic and the normal groups had a mean age of 44 years with a standard deviation of 4 years. Mean years of school completed were 10.5 years for the schizophrenics and 10.6 years for the normals, with standard deviations of 1.9 and 1.7 years, respectively.

Two measures of intellectual functioning were computed for the groups. Present functioning intelligence was estimated by the IQ score obtained from the CVS. As might be expected, the normals received significantly higher IQ scores, with a mean of 116.7 as compared with 90.7 for the schizophrenics. This disparity in present functioning intelligence may reasonably be attributed to the schizophrenic pathology. On the basis of the equivalent age and education, the two groups were assumed to be of equal premorbid intelligence.

In addition, premorbid intelligence was estimated by the item number of the highest vocabulary item passed, a measure introduced by Becker (1956). This would appear to be a better estimate than the more usual measure of total vocabulary score because of the unevenness which characterizes schizophrenic performance. On this measure, the mean score for the schizophrenics was 26.1 and 28.3 for the normals. The schizophrenics were more variable than the normals, with a standard deviation of 7.0 as against 4.6 for the normals. Although a comparison of the scores on the basis of a Mann-Whitney *U* test showed a slight trend toward significantly higher normal scores ($p < .13$), it seems likely that the greater variability of the schizophrenic group is attributable to their pathology, and, further, the pathology would be expected to induce this variability by lowering some scores and raising none, thereby tending to depress somewhat the mean for the schizophrenics. It therefore seemed that the results on this measure are not incompatible with the assumption of equivalent premorbid intelligence in the groups.

TABLE 1
ERROR SCORES ON ANTONYM TASK

	Predicted	Unidentified	Predicted minus Unidentified
Normals			
Mean	1.3	1.1	0.2
Median	0.5	0.9	−0.1
SD	1.7	1.3	2.1
Schizophrenics			
Mean	9.2	3.2	6.0
Median	4.5	1.8	2.3
SD	11.0	3.8	9.1

TABLE 2
ERROR SCORES ON HOMONYM TASK

	Predicted	Unidentified	Predicted minus Unidentified
Normals			
Mean	0.2	0.5	−0.3
Median	0.1	0.3	−0.2
SD	0.6	0.8	0.9
Schizophrenics			
Mean	3.4	1.1	2.2
Median	1.5	1.0	0.8
SD	4.9	1.2	4.6

RESULTS

The means, medians, and standard deviations of the error scores were computed for both groups. These results are summarized for the Antonym and Homonym tasks in Tables 1 and 2, respectively.

Within each group, comparisons of distracter and irrelevant markings were made for each task, using two-tailed Wilcoxon tests. It was found that on both tasks, the schizophrenics chose the predicted error alternative more often than the unidentified error alternatives. On the other hand, the normals showed no significant preference for either type of error.

In order to compare the groups more directly on this differential error preference, there was computed for each subject the difference score of predicted error markings minus unidentified error markings. As discussed above, this is the corrected measure of antonym or homonym choice. On both tasks, the schizophrenics exceeded the normals ($p < .001$). This confirmed the hypothesis that schizophrenics, more than normals,

treat antonyms as synonyms and treat homonyms as synonyms.

In order to study the relationship of intelligence to this feature of thinking disorder, the rank order correlation, corrected for ties, was computed between the CVS score and each of the two difference scores. Although one might expect negative correlations of sizable magnitudes between intelligence measures and this measure of cognitive disturbance, for the normal subjects there appeared to be essentially no correlation, the actual values being .16 for the Antonym task and .18 for the Homonym task. For the schizophrenics, the correlation for both tasks with the CVS IQ was −.44, a significant negative relationship ($p < .05$). These significant values for the patient group probably reflect the fact that both functioning IQ decrement and cognitive disturbance increase with an increase in overall pathology.

In an earlier study by Burstein (1959), the Antonym task was given to third and sixth grade children, and, as predicted, the younger children scored significantly higher than the older on the corrected measure of antonym choices (antonym markings minus unidentified error markings). Since in the present study schizophrenics were distinguished from normals on this same measure, the results support a regression theory of schizophrenic thinking, at least on a descriptive level.

The formulation of the hypothesis in primary process terms does not, of course, preclude other explanations of the data. There exists a considerable body of research with children dating from Razran (1939) in which conditioning techniques have been used to investigate verbal relationships of a sort similar to those considered here. Typically, the approach is to use a visually presented word as the conditioned stimulus and a GSR as the conditioned response. After conditioning is established, homonyms, antonyms, etc. are presented and generalization curves established. These curves are usually found to vary as a function of age, and the gradient of generalization is said to have altered.

One may interpret the greater treatment by schizophrenics of antonyms and homonyms as synomyms as an expression of increased generalization, following interpretations of similar results with children. Mednick (1958) has speculated that schizophrenia in general is interpretable in terms of broadened generalization curves.

SUMMARY

On the basis of the psychoanalytic concept of primary process, it was predicted that schizophrenics would show a greater tendency than normals to treat both antonyms and homonyms as synonyms. The instrument used was a multiple-choice paper-and-pencil test. Each item required the subject to select a synonym from among three alternatives, a synonym, either an antonym or a homonym, and a third word which was neither and was included to measure choices made upon unidentified bases (e.g., obscure communalities, random markings, etc.) which might tend to inflate the number of distracter markings. A corrected score consisting of the difference between predicted error markings and unidentified error markings was therefore taken as a relatively pure measure of the tendency to treat antonyms or homonyms as synonyms, independent of other error tendencies.

The schizophrenics exceeded the normal subjects on this corrected score for both the antonym task and the homonym task.

A possible non-Freudian interpretation of the data in terms of learning theory generalization is discussed.

REFERENCES

BECKER, W. C. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **53**, 229–236.

BURSTEIN, A. G. Primary process in children as a function of age. *J. abnorm. soc. Psychol.*, 1959, **59**, 284–286.

FREUD, S. *Collected papers*. Vol. 4. London: Hogarth, 1925.

FREUD, S. *An outline of psycho-analysis*. New York: Norton, 1949.

FREUD, S. *The interpretation of dreams*. New York: Basic Books, 1956.

HUNT, W. A., FRENCH, ELIZABETH G., KLEBANOFF, S. G., MENSH, I. N., & WILLIAMS, M. Clinical possibilities of an abbreviated individual intelligence scale. *J. consult. Psychol.*, 1948, **12**, 171–173.

MEDNICK, S. A. A learning theory approach to research in schizophrenia. *Psychol. Bull.*, 1958, **55**, 316–327.

RAZRAN, G. H. S. A quantitative study of meaning by a conditioned salivary technique. *Science*, 1939, **90**, 89–90.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Bureau of Publication, 1944.

(Received October 10, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 158-160

SOME INCONSISTENCIES IN THE EFFECT OF MOTIVATION AROUSAL ON HUMOR PREFERENCES[1]

DONN BYRNE[2]
University of Texas

A recent article by Strickland (1959) reported the results of a study in which annoyed and frustrated subjects showed the greatest preference for hostile cartoons while sexually aroused subjects preferred cartoons with a sexual theme. With the Zeitgeist operating, both Strickland (1958) and the present author (Byrne, 1957) attacked an almost identical problem in their doctoral dissertations, but with different results. This paper is designed to point out the similarities and differences between the two studies in the hope of leading to a discovery of the reasons for the inconsistencies. While a large portion of the present study was also concerned with the measurement of defense mechanisms, only the relevant aspects of the experiment will be described.

METHOD

Cartoons. Using a pool of 400 magazine cartoons, five experienced clinical psychologists sorted the humorous material into four content categories (and an unclassifiable group). The final scale was drawn from those cartoons on which there was 80% or better interjudge agreement. It consisted of 16 hostile cartoons (depicting the expression of aggressive drive), 16 sexual (depicting the expression of sexual drive), 16 ridicule (no overt hostility but cartoonist seems to be ridiculing the central character), and 16 nonsense (no drive expression, only incongruity, a play on words, etc.).

The cartoons, in random order, were presented to the subjects in a group situation. Each subject was required to rate the cartoons in four degrees of funniness (most funny to least funny) with 16 cartoons in each category. A cartoon placed in the most funny category received a score of three, in the next most funny a score of two, etc. Each subject received a score on each of the four cartoon types, which could theoretically range from 0 to 48.

Subjects. The subjects consisted of 76 male undergraduates enrolled in an introductory psychology course at Stanford University. The experimental groups (aggression arousal and sex arousal) each consisted of 23 subjects, and the control group was made up of 30 subjects.

Conditions. The method selected to arouse hostile feelings was one employed successfully by Weatherley (1956). Briefly, a second experimenter asked for 5 minutes of the subjects' time for a pilot study in which a confusing "spatial concentration test" was administered in an insulting and provoking manner. Immediately afterward, the cartoon test was given.

The sex arousal group read 10 literary passages selected on the basis of their sexual content (e.g., *God's Little Acre, Ulysses, Aphrodite*). Each subject read these passages (5.5 dittoed pages) and then was asked to rank them in terms of the desirability of the described behavior. When this task was finished, the cartoons were administered.

Half of the control group took the spatial concentration test from a mild-mannered, polite experimenter. The other half read neutral passages from the same books from which the sexual material was taken. These were ranked by the subjects with respect to how much they would like to see the described scene. Again, the cartoon test was given immediately afterward.

RESULTS

In Table 1, the means and standard deviations of the four cartoon scores in the three conditions are given. The effect of the conditions on the humor scores was determined by analysis of variance. For the hostile, sex, and nonsense cartoons, motivation arousal did not exert a significant influence on cartoon ratings. The ridicule cartoons were rated significantly less funny ($p < .01$) in the sex arousal condition. The meaning of this finding, should it be verified in future studies, is unclear.

Within each condition, there was a good deal of consistency in that the hostile cartoons were considered the most funny while the other three types received very similar ratings (with the exception noted above). Were it not for Strickland's findings, it would be concluded that humor preferences are independent of motivational states.

DISCUSSION

Independent experimental results which are not congruent are always a source of embarrassment and concern. It is, of course, possible that motivation arousal does not influence humor preferences and that Strickland unwittingly reported chance findings. Actually, other attempts to relate drive arousal to humor response have yielded uniformly negative results (Byrne, 1958; O'Connell, 1960; Sears, Hovland, & Miller, 1940).

However, the high level of statistical significance reported by Strickland ($p < .005$) makes it

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree at Stanford University.

[2] The author wishes to express his thanks to C. L. Winder, Albert Bandura, Paul R. Farnsworth, and Quinn McNemar for their advice and assistance.

plausible to assume that his findings are real ones; therefore, the reasons for the conflicting results should be sought. Since neither of these two studies was designed as a replication of the other, there are numerous minor differences in method. Even though the populations from which the subjects were drawn and the overall designs of the two experiments were remarkably similar, they differed in several respects as is shown in Table 2.

Any one or a combination of these differences could conceivably account for the divergent results. Perhaps the two most important factors were the differences in the administration of the cartoon test and in the methods used to rate cartoon funniness. It has been found that general ratings of humor differ in group and individual situations (Perl, 1933), and it may be that the influence of motivation arousal is dependent on such conditions as this. It should be noted that none of the negative studies mentioned above used individual administration of the humorous material. Also, the way the cartoons were rated could have been the important difference. In the Strickland study, the scores of a given cartoon category were free to vary independently of one another. In the present study, a given category could not be rated more amusing without causing a corresponding decrease in the rating of another category. It might well be that an independent rating of each cartoon is preferable to having four categories with an equal number of cartoons required in each.

Additional work will be necessary in order to isolate the crucial factors. In any event, the generality of Strickland's findings unfortunately must be restricted until further research is undertaken.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF HUMOR SCORES FOR EACH CONDITION

Conditions	Humor Scores			
	Hostile	Sex	Ridicule	Nonsense
Aggression Arousal				
M	25.74	23.35	23.61	23.83
SD	4.89	7.06	3.55	4.04
Sex Arousal				
M	27.57	23.87	20.76	24.26
SD	4.17	4.67	3.21	4.17
Control				
M	25.61	23.59	23.89	23.80
SD	4.90	6.34	3.20	5.02

TABLE 2

COMPARISON OF THE TWO EXPERIMENTS WITH RESPECT TO METHOD

	Strickland	Byrne
Definition of Cartoon Content	The exact wording of the definitions presented to the judges was not the same.	
Cartoons	The specific cartoons used as stimulus material were not the same.	
Judges	15 faculty members, graduate students, and undergraduates	5 experienced clinical psychologists
Criterion of Cartoon Acceptance	Agreement of all judges	Agreement of four out of five judges
Number of Cartoons	11 each of hostile, sexual, and nonsensical from a pool of 150	16 each of hostile, sexual, ridicule, and nonsense from a pool of 400
Administration of Cartoons	Individual	Group
Rating of Funniness	From not funny to extremely funny on 15-point scale	From least funny to most funny in a 4-point categorization (16 cartoons in each)
Aggression Arousal	Experimenter showed displeasure, left subject alone for 20 minutes, and then gave cartoon test.	Experimenter administered a confusing task in an insulting and provoking manner; then another experimenter gave cartoon test.
Sex Arousal	Subjects viewed photos of nude models.	Subjects read sexual passages from novels.

REFERENCES

BYRNE, D. Response to humor as a function of drive arousal and psychological defenses. Unpublished doctoral dissertation, Stanford University, 1957.

BYRNE, D. Drive level, response to humor, and the cartoon sequence effect. *Psychol. Rep.*, 1958, **4**, 439–442.

O'CONNELL, W. E. The adaptive functions of wit and humor. *J. abnorm. soc. Psychol.*, 1960, **61**, 263–270.

PERL, RUTH E. The influence of a social factor upon the appreciation of humor. *Amer. J. Psychol.*, 1933, **45**, 308–312.

SEARS, R. R., HOVLAND, C. I., & MILLER, N. E. Minor studies of aggression: 1. Measurement of aggressive behavior. *J. Psychol.*, 1940, **9**, 275–295.

STRICKLAND, J. F. The effect of motivation arousal on humor preferences. Unpublished doctoral dissertation, Johns Hopkins University, 1958.

STRICKLAND, J. F. The effect of motivation arousal on humor preferences. *J. abnorm. soc. Psychol.*, 1959, **59**, 278–281.

WEATHERLEY, D. A. An investigation of the relationship between anti-semitism and displacement, projection, and inhibition of aggression. Unpublished doctoral dissertation, Stanford University, 1956.

(Received October 9, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 161-164

THE EFFECT OF INDUCED STRESS ON FREE ASSOCIATIONS[1]

ELI SALTZ

Center for the Study of Cognitive Processes, Wayne State University

EVIDENCE to date is equivocal concerning the effect of stress on such complex cognitive activities as free associations. Clinicians typically assume that stress or anxiety associated with specific concepts will reduce the likelihood of common associations to these concepts. Neurotics and psychotics do tend to give fewer common responses than normals on free association tests (Woodworth & Schlosberg, 1954), and Buchwald (1959) reported that subjects who are anxious on the Taylor scale also give fewer common responses. These results raise the question: Does stress or anxiety reduce the strength of strong associations in complex verbal materials? The Taylor-Spence (1952) position argues that anxiety should *increase* the strength of a dominant response, but results like those of Saltz and Hoehn (1957) have thrown serious doubt on this prediction.

The point of view taken here holds that in the acquisition of any verbal connection, the major problem is not the formation of the individual association. It is the differentiation of the given association from all other associations. This is obvious in paired associates learning. Any single pair can usually be learned in one trial. Difficulty in learning is a consequence of learning many pairs in a single list. Reinforcement of a given pair can be conceptualized as strengthening the boundary strength between this pair and other pairs, thus producing differentiation. Stress acts to break down these boundaries and to produce dedifferentiation. Thus, stress reduces the frequency of common responses in free associations.

In the present study, stress was induced in a manner similar to that suggested by Beam (1955). Beam had college subjects learn material immediately prior to course examinations and found significant decrements in learning. Preliminary to the present study, subjective reports were obtained from students concerning their "felt anxiety" in examination situations. More students indicated anxiety *after* taking a difficult exam and immediately before receiving their exam grades than reported anxiety immediately *before* taking an exam. Consequently, the stress condition consisted of administering the free association test immediately *after* the first full exam of the course. The instructor informed the subjects that he would grade their papers while they were taking the free association test and would give them their grades at the end of the hour when all students had completed the experiment. The course exam was a difficult one, the highest grade being 49 correct items out of 60.

[1] The present study is part of a program of research, supported by a grant to the writer from the National Science Foundation, on the influence of differentiation on verbal processes.

METHOD

Subjects. Fifty-seven Wayne State University students, enrolled in summer classes in introductory psychology, served as subjects.

Material and Procedure. The 100-item Kent-Rosanoff free association list was used. Instructions and procedure for administration were identical to those described by Russell and Jenkins (1954): The 100 items were typed on two pages, two columns of 25 stimuli on each page, with a blank line for subject to write his association following each stimulus word. Subjects were told to write the first word that came to mind upon seeing each stimulus word. There was no time limit.

Subjects took the test twice, the administrations separated by one week, once under no stress and once under stress. Thirty subjects had the nonstress condition first; these were tested in two subgroups, one of 12 subjects and the other of 18 subjects. Twenty-seven subjects had the stress condition first; these were tested in two subgroups, one of 13 subjects, the other of 14 subjects. Since the results for subgroups within each order of presentation were similar, the subgroups within each order were combined for purposes of analysis.

Analysis. Only those stimuli were used for which some single response was clearly dominant in the Russell and Jenkins restandardization of the Kent-Rosanoff. Items were divided into three categories on the basis of the frequency with which the dominant response to each of the selected stimuli occurred in the Jenkins study: 20% to 40% occurrence (21 words), 41% to 60% (17 words), 61% and above (19 words). The words in each category are indicated in Table 1. For each subject, the percentage of common responses in each category was computed. These data were analyzed by means of analysis of variance for related measures.

RESULTS

As can be seen in Figure 1, the effect of stress was to reduce the number of common associations elicited during free associations. This result occurred both when the stress condition preceded the nonstress and when the stress trials followed the nonstress. The reduction in common associations as a function of stress is significant at beyond the .001 level.

TABLE 1

STIMULUS WORDS, AND THEIR DOMINANT ASSOCIATIONS, USED IN THE PRESENT STUDY

Frequency of Occurrence of Dominant Response[a]		
20%–40% Stimulus—Response	41%–60% Stimulus—Response	61% and over Stimulus—Response
1. Sickness—health (38%)	Soft—hard (44%)	Table—chair (84%)
2. Deep—shallow (32%)	Chair—table (49%)	Dark—Light (83%)
3. Eating—food (39%)	Sweet—sour (43%)	Man—woman (77%)
4. Mountain—hills (27%)	Rough—smooth (44%)	Black—white (75%)
5. House—home (25%)	Spider—web (45%)	Woman—man (65%)
6. Fruit—apple (38%)	Needle—thread (46%)	White—black (62%)
7. Smooth—rough (33%)	Carpet—rug (46%)	Slow—fast (75%)
8. Command—order (20%)	Sour—sweet (57%)	Girl—boy (70%)
9. Cold—hot (35%)	Eagle—bird (55%)	High—low (68%)
10. Sleep—bed (24%)	Stem—flower (40%)	Hard—soft (67%)
11. Anger—mad (35%)	Dream—sleep (45%)	Lamp—light (63%)
12. Bath—clean (31%)	Hammer—nails (54%)	Bread—butter (61%)
13. Cottage—house (30%)	Loud—soft (54%)	Boy—girl (77%)
14. Swift—fast (37%)	Bed—sleep (58%)	Light—dark (65%)
15. Hungry—food (36%)	Heavy—light (58%)	Long—short (76%)
16. Priest—church (33%)	Tobacco—smoke (56%)	Butter—bread (64%)
17. Square—round (37%)	Salt—pepper (43%)	Scissors—cut (67%)
18. Doctor—nurse (24%)		King—queen (75%)
19. Thief—steal (29%)		Blossom—flower (67%)
20. Lion—tiger (26%)		
21. Quiet—loud (35%)		

[a] The percentages indicate the frequency with which the dominant response to each stimulus was elicited in the Russell and Jenkins (1954) standardization.

The number of common associations given during nonstress is almost identical in the two orders of presentation, as can be seen in Figure 1. While there appears to be a tendency for stress to produce somewhat less of a decrement when the stress condition precedes the nonstress, this difference does not approach the .05 level, suggesting that order of presentation may not be a critical variable.

Association level, on the other hand, was a significant variable here in determining the drop in frequency of common responses as a function of stress (F significant at beyond the .05 level). From Figure 1 it is apparent that the smallest decrements occur to the low association value items. For the nonstress-stress order of presentation, there is little difference between high association value items and middle association items. For high association value items, the drop was 10.8 percentage points. On the other hand, the stress-nonstress order does show some tendency for a greater drop in the middle association value items (7.0 percentage points drop) than by the high association value items (3.7 percentage points drop). Since neither the latter difference nor the difference between orders is significant, the safest induction for the present is that there is no real difference between high and medium association values in the decrement produced by stress.

The writer analyzed the associations to determine if stress tended to produce a tendency to give responses which were outside of the total list of responses reported by Russell and Jenkins (1954). No such tendency was observable. Under both stress and nonstress conditions, failure to give the common response usually resulted in giving some other word in the Russell and Jenkins list for the particular stimulus.

DISCUSSION

The results of the present study are consistent with Buchwald's (1959) findings with anxiety (high and low scores on the Taylor scale). They suggest that stress and anxiety reduce the strength of common responses in free associations, contrary to what is anticipated on the basis of a drive theory of stress such as proposed by Castaneda and Palermo (1955). It might be pointed out that in their studies of the effect of stress on discrimination learning (Castaneda, 1956; Castaneda & Palermo, 1955; Palermo, 1957), these writers found no significant effects upon well-learned discriminations as a result of stress. This may be a consequence of the type of stress used in these studies: Castaneda and Palermo attempted to

FIG. 1. Mean percentages of common responses in free associations as a function of stress, order of presentation, and strength of association.

induce stress by telling their subjects to work faster, or by excessive pacing of the learning trials.

Similarly the paucity of novel responses found among the free associations during stress suggests that the type of explanation proposed by Child (1954) to account for anxiety scale results is not applicable here. Child suggests that anxiety has unique stimulus characteristics which produce responses which interfere with the responses to the stimulus items presented by the experimenter. Other methods of producing stress (e.g., electric shock) may have more characteristic response associates. The present study argues that such response associates are not essential for producing decrements in the common responses.

A simple response competition theory (modeled after the theory proposed by Child to account for anxiety scale results) would lead to the prediction that stress has a linear effect on the reduction of common response. This is true because the amount of stress induced response competition is constant, no matter what the strength of the word associations involved at any given moment.

Farber (1954) has suggested a two-factor position, containing both the drive notion and Child's notion concerning the stimulus characteristics of stress. As a stimulus, stress produces characteristic stress responses which may compete with the common free associations. As a drive, stress multiplies the strength of both the common free associations and their competing stress responses. Consequently, the drive characteristic of stress should strengthen the strong common free associations disproportionately more than the weak common associations. Thus, the strong common associations should suffer less of a decrement than the weak common associations. Just the opposite occurred.

The results of the interaction between stress and association strength of the common responses suggests that dedifferentiation is a curvilinear function of boundary strength. This notion seems best to explain the phenomena under consideration.

SUMMARY

The purpose of the present paper was to investigate the influence of stress on the common responses in free associations. On the basis of a dedifferentiation theory, it was predicted that stress would reduce the frequency of common responses. On the basis of a drive theory, one might anticipate an increase in frequency of common responses.

Stress was operationally defined as a situation in which subjects had just taken a difficult course examination and were waiting to be told their grades.

Results indicated highly significant decrements in common responses as a function of stress. There was no tendency toward "novel" responses (responses not given to a stimulus in the Russell and Jenkins standardization). The theoretical implications of these observations were discussed.

REFERENCES

BEAM, J. C. Serial learning and conditioning under real life stress. *J. abnorm. soc. Psychol.*, 1955, **51**, 543–552.

BUCHWALD, A. M. Manifest anxiety-level, verbal response-strength, and paired-associate learning. *Amer. J. Psychol.*, 1959, **72**, 89–93.

CASTANEDA, A. Effect of stress on complex learning and performance. *J. exp. Psychol.*, 1956, **52**, 9–12.

CASTANEDA, A., & PALERMO, D. S. Psychomotor performance as a function of amount of training and stress. *J. exp. Psychol.*, 1955, **50**, 175–179.

CHILD, I. L. Personality. *Annu. Rev. Psychol.*, 1954, **5**, 149–170.

FARBER, I. E. Anxiety as a drive state. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1954.* Lincoln: Univer. Nebraska Press, 1954.

PALERMO, D. S. Proactive interference and facilitation as a function of amount of training and stress. *J. exp. Psychol.*, 1957, **53**, 293–296.

RUSSELL, W. A., & JENKINS, J. J. The complete

Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. *USN Off. Naval Res. tech. Rep.*, 1954, No. 11. (Contract No. N8)

SALTZ, E., & HOEHN, A. J. Test of the Taylor-Spence theory of anxiety. *J. abnorm. soc. Psychol.*, 1957, **54**, 114–117.

TAYLOR, J. A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **42**, 61–64.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. (Rev. ed.) New York: Holt, 1954.

(Received October 9, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 165–168

THE EFFECTS OF ANXIETY AND THREAT ON THE SOLUTION OF A DIFFICULT TASK[1]

IRWIN G. SARASON[2]
University of Washington

This experiment was concerned with the performance of subjects differing in anxiety scale scores on a concept formation task known to be a very difficult one. The subjects differed not only with respect to anxiety scores but also with respect to the instructions administered to them prior to performance. Underlying the research was the hypothesis that instructions and degree of anxiety interact in affecting performance. In view of previous research, it seemed that an adequate test of this hypothesis could best be made using a task high in difficulty level (Nicholson, 1958; Sarason & Palola, 1960).

There are at least two important aspects of a difficult or complex task. Obviously, a complex task is hard, and successful performance on it requires intellectual ability commensurate with the level of the problem. But there is another aspect of a highly complex task—its potential for arousing emotional reactions in some subjects. Several studies using anxiety scales have reported poorer performance for high anxious than low anxious subjects on complex or difficult tasks (Montague, 1953; Ramond, 1953; Taylor & Spence, 1952). Other investigators, however, have not found significant differences due to anxiety on tasks which appear to be relatively high in difficulty level (Axelrod, Cowen, & Heilizer, 1956; Saltz & Hoehn, 1957). It would appear that the usefulness of present conceptions of anxiety in predicting subjects' reactions to challenging, difficult tests of intellectual ability is as yet quite limited.

Better assessment of individuals' reactions to a difficult intellectual problem may be developed if, in addition to the difficulty and anxiety variables, the factor of the conditions under which performance takes place is better understood. One important aspect of these conditions is the nature of the instructions administered to subjects immediately prior to performance. In the present study, prior to subjects' attempts at solving 13 very difficult anagrams, half the subjects were told that the anagrams could be easily solved by individuals of average intelligence. These instructions were called the personal threat condition because all subjects quickly discovered that they could not solve all of the anagrams. Under this condition, it was predicted that low anxious subjects would be superior to high anxious subjects. The prediction is based on the assumption that high anxious subjects respond to a condition of personal threat with more personalized anxiety responses than do low anxious subjects.

The remaining subjects were told, prior to performance, that the anagrams were actually quite difficult and that successful solution of all of them in the allotted time was highly unlikely. It was predicted that under this condition low anxious subjects would not perform at a significantly higher level than would high anxious subjects. This was expected on the assumption that this set of instructions would reassure high anxious subjects and reduce their anxiety over poor performance.

An additional aspect of the present study concerns the comparison of three anxiety scales with respect to their usefulness in assessing subjects' potential for emission of interfering responses under the personal threat condition. One of the scales, the Test Anxiety Scale (Sarason, 1958b), dealing specifically with subjects' reactions to testing situations, was expected to interact more significantly with instructions than would two indices of general anxiety employed.

METHOD

Subjects

The subjects were 191 students in introductory psychology courses at the University of Washington. Independent of and prior to participation in the experiment proper, these subjects were administered the Test Anxiety Scale (TAS) and Lack of Protection Scale (LP) of the Autobiographical Survey (Sarason, 1958a, 1958b). Ninety-one of the subjects were also administered the Bendig short form of the Taylor Manifest Anxiety Scale (MAS) (Bendig, 1956). Since the personality test data and the performance scores of the subjects were obtained independently of each other, certain of the subjects had to be randomly omitted from each of the statistical analyses in order to meet the analysis of variance requirement of proportionality of cell frequencies.

On TAS, the subjects were divided into high, middle, and low anxious groups. The extreme groups corresponded approximately to the upper and lower 30% of the TAS score distribution. The middle TAS

[1] This research was supported by a grant from the National Institute of Mental Health (M2397).

[2] The writer is indebted to William R. Crow and James G. Minard for their assistance in carrying out this study.

TABLE 1
THIRTEEN ANAGRAMS ADMINISTERED TO ALL SUBJECTS

1. ETLHHA	8. SUTCBII
2. ETROS	9. SCLIAO
3. RECMI	10. EVSUORN
4. CNEGAH	11. RSANEO
5. NMGOINR	12. IMTCELA
6. NSRWAE	13. ELSAUX
7. SPRUUE	

group corresponded to the middle 40% of the distribution. High TAS subjects had scores of 6–16; middle TAS subjects had scores of 3–5; and low TAS subjects had scores of 0–2.

On LP the subjects were also divided into three groups corresponding to the upper and lower 30% of the distribution and the middle 40%. High LP scores ranged between 9–20; middle LP scores ranged between 4–8; and low LP scores ranged between 0–3.

Because of the smaller MAS N, the subjects were divided at the median on this scale. High MAS subjects had a range of 6–17, and low MAS had a range of 0–5.

Procedure

The anagrams used were selected from Sargent's list of those of high difficulty (Sargent, 1940). Table 1 gives the anagrams employed.

All subjects were tested in groups ranging in size from 35–50. Each subject was given a booklet consisting of two pages with the instructions on the first page and the list of anagrams to be solved on the second. The time limit for all subjects was 18 minutes.

The instructions page consisted of two paragraphs. For all subjects the first paragraph provided the following information:

> On the next page you will see a series of disarranged words. Your job will be to rearrange each group of letters so that they make a meaningful English word. Start when you are so instructed. Stop at the stop signal. Write your name at the top of the next page when given the signal.

The subjects under the personal threat condition were given booklets which had the following second paragraph on the first page:

> Ability to organize material such as the letters on the next page has been found to be directly related to intelligence level. High school students of above average intelligence (IQ greater than 100) and most college students should be able successfully to complete the task. You will have 18 minutes in which to complete it.

The subjects under the nonthreat conditions were given booklets which had the following second paragraph on the first page:

> Most of you probably have worked anagrams. The task on the next page works the same way. These anagrams, however, are harder than most you have seen in books and magazines. Consequently, you may not finish all of them and you may find some of the anagrams very difficult. If this happens, don't worry about it. No one will find the anagrams easy.

The seating arrangement of the testing room was such that none of the subjects could have known that there were two forms of the instructions.

RESULTS

A 2 × 3 analysis of variance was performed on the number of correct anagram solutions with three levels of TAS (high, middle, and low) and two levels of instruction (threat and nonthreat). Only one F was statistically significant, that for the interaction between test anxiety and instruction ($p < .01$). Under the threat condition, the middle and low TAS subjects performed at a higher level than did the high TAS subjects. However, just the opposite was the case under the nonthreat condition, where the high TAS subjects were superior to middle and low subjects. Table 2 presents the means and standard deviations for the six experimental groups.

A similar 2 × 3 analysis for LP also revealed only one significant effect, that for the anxiety by instructions interaction ($p < .025$). The nature of this interaction was identical to the one obtained when subjects were divided on the basis of TAS. Table 3 presents the means and standard deviations for the LP analysis.

A 2 × 2 analysis of variance carried out on the MAS data failed to reveal any significant effects.

TABLE 2
MEANS AND STANDARD DEVIATIONS FOR NUMBER OF CORRECT SOLUTIONS FOR SIX EXPERIMENTAL GROUPS DIVIDED ON THE BASIS OF TEST ANXIETY SCALE (TAS) AND INSTRUCTIONS

	Threat Instructions			Nonthreat Instructions		
TAS	*M*	*SD*	*N*	*M*	*SD*	*N*
High	3.54	2.13	26	5.31	3.22	26
Middle	4.73	2.64	37	3.49	2.15	37
Low	4.92	2.61	25	4.04	3.17	25

TABLE 3
MEANS AND STANDARD DEVIATIONS FOR NUMBER OF CORRECT SOLUTIONS FOR SIX EXPERIMENTAL GROUPS DIVIDED ON THE BASIS OF LACK OF PROTECTION SCALE (LP) AND INSTRUCTIONS

	Threat Instructions			Nonthreat Instructions		
LP	*M*	*SD*	*N*	*M*	*SD*	*N*
High	3.93	2.33	29	5.38	2.88	29
Middle	4.54	2.70	37	3.92	2.61	37
Low	4.90	2.50	21	3.48	2.40	21

although the anxiety by instructions interaction was in the same direction as the significant interactions reported above.

DISCUSSION

The empirical findings of the present research are quite clear-cut. For two of the three anxiety scales used, it was found that, according to the predictions made, threat and nonthreat conditions differentially affected high, middle, and low anxious subjects. Under the threat conditions, the performance of high anxious subjects was on a significantly lower level than the performance of subjects in the rest of the anxiety score distribution. Just the opposite result emerged under the nonthreat condition. Although the anagrams task was clearly a very difficult one and, judging from informal comments of subjects after the experiment, a frustrating one also, it would seem that the nonthreat instructions may have functioned as a reassurance to high anxious subjects.

Studies by Sarason and Palola (1960) and Nicholson (1958) have provided evidence suggesting that the optimal conditions for arousal of interfering responses in high anxious subjects are a combination of high difficulty level and personal threat. Although the difficulty variable was not manipulated in the present experiment, the results are consistent with this generalization. The finding here of a superiority of high to middle and low anxious subjects under the nonthreat condition seems to be a potentially important one. It is possible that a condition which provides reassurance and consequently better performance for a high anxious individual may actually serve to lower the motivational level and performance of less anxious individuals (Sarason, 1958a).

If these inferences are extended to the general problem of intellectual evaluation, it may well be that when intelligence or aptitude testing is conducted under conditions of personal threat (as perceived by the subject), high anxious subjects' ability will be systematically underestimated due to their emission of interfering anxiety responses. This possibility would seem particularly likely in the case of college entrance examinations, where the student is well aware of the evaluative and selective purposes of testing. Certainly, the present study, together with previous research, suggests the need for more intensive study of the role of individual difference variables such as anxiety in intellectual performance (Sarason, 1957, 1959a).

It had been expected on the basis of previous research that the test anxiety measure would more significantly interact with instructions than would the LP and MAS measures (Sarason, 1958a, 1959b, 1959c). However, significant interactions were found for both TAS and LP. It would seem that there are serious shortcomings in present knowledge of (*a*) the conditions under which only the specific test anxiety index relates to performance, and (*b*) the conditions under which both general and specific indices of anxiety relate to performance.

SUMMARY

Two variables were manipulated in this research, anxiety and experimental instructions administered prior to the subjects' performance on a series of anagrams high in difficulty level. Three measures of anxiety were used: the Test Anxiety Scale (TAS), the Lack of Protection Scale (LP), and Bendig's version of the Manifest Anxiety Scale (MAS). The two sets of instructions differed in that one set was designed to pose a personal threat to the subject and the other set was designed to reassure the subject.

The results showed that TAS and LP, but not MAS, significantly interacted with the instructions. High anxious subjects under the threat condition performed at a lower level than did low and middle anxious subjects. Just the reverse was true under the nonthreat conditions, where high anxious subjects were superior to low and middle anxious subjects.

The results were interpreted in terms of the interfering responses aroused in high anxious individuals under personal threat conditions. The problem of the specificity of anxiety scales was also discussed.

REFERENCES

AXELROD, H. D., COWEN, E. L., & HEILIZER, F. The correlates of manifest anxiety in stylus maze learning. *J. exp. Psychol.*, 1956, **51**, 131–138.

BENDIG, A. W. The development of a short form of the manifest anxiety scale. *J. consult. Psychol.*, 1956, **20**, 384.

MONTAGUE, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–95.

NICHOLSON, W. M. The influence of anxiety upon learning. *J. Pers.*, 1958, **26**, 303–319.

RAMOND, C. Anxiety and task as determiner of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120–124.

SALTZ, E., & HOEHN, A. J. A test of the Taylor-Spence theory of anxiety. *J. abnorm. soc. Psychol.*, 1957, **54**, 114–117.

SARASON, I. G. Test anxiety, general anxiety, and intellectual performance. *J. consult. Psychol.*, 1957, **21**, 485–490.

SARASON, I. G. Effects on verbal learning of anxiety, reassurance, and meaningfulness of material. *J. exp. Psychol.*, 1958, **56**, 472–477. (a)

SARASON, I. G. Interrelationships among individual difference variables, behavior in psychotherapy, and verbal conditioning. *J. abnorm. soc. Psychol.*, 1958, **56,** 339–344. (b)

SARASON, I. G. Intellectual and personality correlates of text anxiety. *J. abnorm. soc. Psychol.*, 1959, **59,** 272–275. (a)

SARASON, I. G. Relationships of measures of anxiety and experimental instructions to word association test performance. *J. abnorm. soc. Psychol.*, 1959, **59,** 37–42. (b)

SARASON, I. G. Relationships of measures of anxiety and experimental instructions to word association test performance: Further findings. *Amer. Psychologist*, 1959, **14,** 370. (Abstract) (c)

SARASON, I. G., & PALOLA, E. G. The relationship of test and general anxiety, difficulty of task, and experimental instructions to performance. *J. exp. Psychol.*, 1960, **59,** 185–191.

SARGENT, S. S. Thinking processes at various levels of difficulty. *Arch. Psychol., NY*, 1940, No. 249.

TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44,** 61–64.

(Received November 6, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 169–171

EGO STRENGTH AND PERCEIVED COMPETENCE AS CONFORMITY VARIABLES[1]

KAY H. SMITH[2]
Wayne State University

PRESSURE toward conformity is generated whenever a person, as a member of a small group, finds his perceptions or judgments at variance with a majority of the other members of the group. The amount of yielding to such pressure logically should vary with the value placed by the deviate on the ability of the other group members to judge correctly in the situation relative to his perception of his own ability to do so. The two studies available in the literature on this point used ambiguous material. Hochbaum (1954) used as a task the prediction of the outcome of cases from case materials. Kelman (1950) found similar results using the autokinetic phenomenon. The present study is an attempt to validate these findings employing a situation similar to that used by Asch (1952), in which it is rather obvious to the deviate that the other group members are giving erroneous judgments.

Resistance by the deviate to such majority pressure would seem to be dependent upon his capacity to assimilate discrepancies in informational input without being threatened. A general name given to this attribute is "ego strength." Barron (1953, 1954) proposes his *Es* scale as a measure of this variable. Daniels (1957) found that an index using the *Es* scale and the Self-Contextual-Influence test was related to susceptibility to persuasive communication. Crutchfield (1955), using an Asch-type situation (1952), found a correlation of −.33 between conformity and *Es* scores.

The hypotheses to be tested are:

Hypothesis I. Persons who place higher value on their own ability to judge yield less under pressure to conformity than those who place a lower value on their own ability relative to that of the other group members.

Hypothesis II. Persons who have high ego strength yield less under pressure to conformity than persons with low ego strength.

METHOD

Es scale. This scale consists of 68 items taken from the MMPI. The specific items and the method of scoring are found in Barron (1953, 1954). The scale was administered without any filler items. The items were arranged in the order in which they appear in the long form of the MMPI.

Judgment apparatus. Judgments were made on the standard Galton bar, with the addition of a sliding clamp fitted with a thumb screw that could be tightened or loosened. The clamp was so arranged that it was hidden from the sight of anyone seated in front of the apparatus, and enabled the confederates in the experiment to make accurately any judgment desired by the experimenter. Each time a confederate made his judgment he would simply adjust the bar until its movement was arrested by the clamp. The clamp was loosened when the critical subject was adjusting the bar to allow freedom of movement in any direction. The side of the bar used as the standard of judgment was set at 30 centimeters.

Subjects. The subjects were 48 students from three introductory psychology classes at Brigham Young University. The *Es* scale was administered in class, the total number of students tested being 253.

Since Daniels (1957) had reported a sex difference in ego strength scores, the means and standard deviations of *Es* scores were computed separately for males and females. For the males the mean was 46.90 and the *SD* 4.55, and for the females the mean was 43.25 and the *SD* 3.68, the means being significantly different at well beyond the .01 level. To take into account this sex difference 12 males and 12 females were counted off from each extreme of the distribution. These subjects comprised the sample for the remainder of the experiment.

Ten male and 12 female Brigham Young University students served as confederates. With the exception of 2 male graduate students all were undergraduates.

Procedure. Each of the selected subjects was taken individually into a room containing the Galton bar and given three trials, each beginning at zero, at adjusting the apparatus so that the sides appeared equal. This yielded the individual average judgment.

At this point in the experiment the subjects were assigned at random to the three experimental groups, four high males, four low males, four high females, and four low females to each group, according to the original ego strength score dichotomy. The procedure in the final testing period for these different experimental groups differed only in the instructions given prior to taking the actual measurement.

The subjects were contacted again and a final testing appointment was set up during which, they were told, they would work with two other subjects. This final session was conducted in the same room as the previous one. One male and one female confederate participated in all cases along with one subject—respectively, either male or female. These three people were introduced to each other as subjects and by name, but beyond this, verbal interaction was kept to a minimum. Each group was told that the group had been assembled to assess

[1] The data reported here are taken from a thesis submitted by the author in partial fulfillment of the requirements for the Master of Arts degree at Brigham Young University.

[2] Acknowledgment is made to Kenneth Hardy whose advice and guidance on this project were greatly appreciated. Gratitude is also expressed to Ross Stagner for his helpful suggestions in the preparation of this manuscript.

the difference between individual judgments and judgments arrived at in cooperation with others. Next the three group members were given percentile scores supposedly indicating how they compared in accuracy with all the people participating in the experiment on the original judgment. For Group A (high self-valuation) the subject was always given a percentile score of 95, the male confederate one of 73, and the female confederate one of 54. For Group B (low self-valuation) the subject was always given the percentile score of 54, the female confederate the score of 73, and the male confederate the score of 95. For Group C (no differential in self-valuation) all three were given the percentile score of 73. These percentile rankings were given orally before the group and the meaning of the score was explained.

The procedure for the group judgment was then explained. The first person was to begin from zero and adjust the bar until he thought the sides were equal, then the second would make whatever adjustments he felt necessary from this point until he was satisfied with the judgment, then the third would do likewise. This final adjustment, they were told, would be taken as the group measure. This procedure was to be followed three times with a different order each time so that each would have an opportunity to make the final adjustment. The three series followed this order: male confederate, subject, female confederate; subject, female confederate, male confederate; female confederate, male confederate, subject. The final measurement of the subject in Series 3 was taken as the critical group measure, since at this point the greatest cumulative effect of the discrepancy should have been achieved between his perception of equality and that demonstrated by the other group members. The difference between this measure and the subject's individual mean measure was taken as the conformity score.

After the session was completed the subject was informed of the purposes of the experiment and of the deception. He was also asked if at any time he had suspicions as to the authenticity of the instructions given, or of any aspect of the situation. In all cases the deception seemed to be complete.

RESULTS

The two hypotheses were tested by means of an analysis of variance. The summary table is given in Table 1, which indicates that the first hypothesis may be accepted as tenable. The second, however, cannot be accepted on the basis of these results.

To test the directionality of difference on the ranking variable, t tests were run between the three experimental groups. The differences between means of the three experimental groups were all in the predicted direction. The difference between the high (A) and low (B) ranked groups ($p < .01$) and the equal (C) and low (B) ranked groups ($p < .05$) were significant. The difference between the high (A) and equal (C) ranked groups approached but did not quite reach the level of significance ($p > .05$). Upon this evidence it seems logical to consider the directionality of the relative ability value hypothesis supported.

TABLE 1
ANALYSIS OF VARIANCE FOR CONFORMITY SCORES

Source of Variation	*SS*	*df*	*MS*	*F*	*p*
Ego strength	.73	1	.73	1.53	>.20
Experimental Groups (ranking)	14.60	2	7.30	6.50	<.01
Interaction	1.47	2	.73	1.53	>.20
Within cells	47.04	42	1.12		
Total	63.84	47			

DISCUSSION

The results obtained here on the relationship of *Es* scores and yielding to group pressure seem to be in opposition to the findings of Daniels (1957) and Crutchfield (1955). With the Daniels work the discrepancy might possibly be due to the use of another measure in addition to the *Es* scale. The major difference between the present study and that of Crutchfield seems to be in the sample used. Since *Es* scores are correlated with intelligence, and since Crutchfield used Air Force officers, there is a possibility that his sample was biased toward the high end of the scale. The present sample included groups from each extreme. A possible consequence of such a difference in sample comes from the work of Janis (1954). Janis reported that while those of his subjects who evidenced the most change seemed to be passive and dependent, those who resisted change deviated from the norm in the direction of neurosis, compulsion, and hypochondriasis. While those scoring low on the *Es* scale have been described as passive and dependent, the scale also contains items indicating possible neurosis, compulsion, and hypochondriasis. Perhaps indeed *Es* scores bear a curvilinear relationship with susceptibility to pressure toward change of judgment.

The results of this study support fully the conclusions reached by Hochbaum (1954) and Kelman (1950) regarding the way in which perceived competence is related to susceptibility to pressure toward conformity. Their conclusions can now be extended to instances in which pressure favors a judgment that is in direct opposition to sensory cues.

SUMMARY

The problem investigated in this research concerned the relationship of two personal variables to susceptibility to group pressure to change. The personal variables used were ego strength, as

measured by the Barron *Es* scale, and perceived competence for the task. Subjects were 12 males, and 12 females selected from each extreme of the distribution of *Es* scores. These subjects were given three trials each on the Galton bar to establish an individual mean. They were then assigned at random to groups and given instruction to establish expectancies of high, low, or average competence on the task. In a group situation confederates gave judgments four centimeters longer than the subject's previous mean. The difference between the subject's final measure in this group situation and his individual mean was his conformity score.

It was concluded that:

1. Subjects selected from the extremes of a distribution of *Es* scale scores do not differ in their susceptibility to group pressure to change.

2. The degree of susceptibility to group pressure to change is related to the subject's own perceived competence for the task involved.

REFERENCES

ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.

BARRON, F. An ego strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333.

BARRON, F. A correction. *J. consult. Psychol.*, 1954, **18**, 150.

CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 190–198.

DANIELS, P. B. The relationship between ego strength and susceptibility to persuasion. Unpublished master's thesis, Brigham Young University, 1957.

HOCHBAUM, G. The relation between group members self-confidence and their reactions to group pressures to conformity. *Amer. sociol. Rev.*, 1954, **19**, 678–687.

JANIS, I. L. Personality correlates of susceptibility to persuasion. *J. Pers.*, 1954, **22**, 504–518.

KELMAN, H. C. Effects of success and failure on "suggestibility" in the autokinetic situation. *J. abnorm. soc. Psychol.*, 1950, **45**, 267–285.

(Received November 12, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 172

THE "YEASAYER" AND SOCIAL DESIRABILITY: A COMMENT ON THE COUCH AND KENISTON PAPER

JAMES B. TAYLOR

Washington Public Opinion Laboratory, University of Washington

A recent paper in this journal investigates the influences of an "agreeing response set" on a number of scales, and explores the relationship of this set to other personality factors. A wealth of data is presented, all of it seeming to demonstrate "both the far-reaching importance of response set in the area of psychological tests and the major proposition that the agreeing response tendency is based on a central personality syndrome" (Couch & Keniston, 1960, p. 173).

There is a possibility, however, that these many findings can also be explained as reflections of the social desirability variable. Couch and Keniston's results may be viewed as once again demonstrating the importance of social desirability in personality scaling, and as giving further information on the "social desirability set"—i.e., the set to endorse socially undesirable statements.

It will be recalled that Couch and Keniston devised a scale to measure agreeing response set. The scale consisted of 360 items drawn from a variety of personality questionnaires. Its items were chosen so as to minimize the influence of specific content. Subjects rated each item on a seven-point scale and were assigned an Over-all Agreement Score (OAS). The OAS had an adequate split-half reliability (.85) and was fairly consistent over 1 or 2 weeks' time.

Now it seems possible that the OAS does not measure an agreeing response set per se, but rather measures the tendency to endorse *socially undesirable* items. The OAS is based on statements drawn from a number of Likert-type scales. Previous research has demonstrated that the majority of such statements can be located on a social desirability continuum. Further, there is a strong tendency for personality scales to be loaded with socially undesirable items. There is also considerable evidence to show that people respond differentially to items on the basis of social desirability: the more socially desirable an item is the more likely it is to be endorsed.

Willingness to endorse socially undesirable items seems to be a general response trait that is manifested irrespective of the obvious content of the item (Edwards, 1957). By choosing a heterogeneous group of items, there is a good chance that Couch and Keniston developed an Over-all Agreement Scale largely made up of socially undesirable items. If this happened, the OAS should be regarded as another measure of the social desirability set—a measure similar to the *SD* scales reported in several other papers (Edwards, 1953; Fordyce, 1956).

The issue seems worth raising, not only for this study but also for other studies that ostensibly measure the agreeing response set. If such studies are measuring the social desirability set as well as (or instead of) the agreeing response set, then it may be that their conclusions need re-evaluation. Perhaps future studies should make an effort to differentiate between the two response tendencies.

REFERENCES

COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, **60**, 151–174.

EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule.* New York: Psychological Corporation, 1953.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

FORDYCE, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, **20**, 171–175.

(Early publication received April 14, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 173-174

A NOTE ON THE COUCH AND KENISTON MEASURE OF AGREEMENT RESPONSE SET

ALLEN L. EDWARDS AND JERALD N. WALKER
University of Washington

EDWARDS (1953) has shown that there is a high positive correlation between the social desirability scale value of a personality item and the probability of endorsement of the item. This finding has been confirmed in a number of other investigations (Cowen & Tongas, 1959; Edwards, 1955, 1957a, 1959; Hanley, 1956; Hillmer, 1958; Kenny, 1956; Taylor, 1959; Wiggins & Rumrill, 1959).

Further consideration of the relationship between the social desirability scale value and the probability of endorsement of a personality item has resulted in the concept of a socially desirable response to a personality item. A socially desirable response is defined as a True response to a personality item with a socially desirable scale value or a False response to a personality item with a socially undesirable scale value (Edwards, 1957b). A 39-item Social Desirability scale (*SD*) has been designed to measure the tendency to give socially desirable responses to personality items and is described by Edwards (1957b). Scores on the *SD* have been found to correlate substantially with scores on many other personality scales of the True-False kind (Edwards, 1957b; Edwards, Heathers, & Fordyce, 1960). In particular, it has been found that if a scale contains a large percentage of items keyed for socially desirable responses, then scores on this scale will correlate positively with scores on the *SD*, regardless of the True-False keying of the items in the scale. Similarly, it has been found that if a scale contains a small percentage of items keyed for socially desirable responses, then scores on this scale will correlate negatively with scores on the *SD*, again, regardless of the True-False keying of the items in the scale.

In a recent article, Couch and Keniston (1960) obtained Over-all Agreement Scores (OAS) on 360 items. They describe the OAS as measuring the tendency of subjects to agree or respond True to items regardless of the content of the items. Part of the evidence reported by Couch and Keniston with respect to the construct validity of the OAS consisted of the correlations of the OAS with 32 Minnesota Multiphasic Personality Inventory (MMPI) scales. They found that MMPI scales which had a large percentage of items keyed True tended to correlate positively with the OAS while MMPI scales with a large percentage of items keyed False tended to correlate negatively with the OAS. They interpret these findings as "... indicating that scales with an unusually high (or low) percentage of items scored True are significantly influenced by the agreeing response set" (p. 154) which they regard the OAS as primarily measuring.

An alternative hypothesis, which Couch and Keniston did not consider, is that scores on both the OAS and on the MMPI scales may also be influenced by the tendency of subjects to give socially *un*desirable responses.[1] The present study is concerned with the evidence concerning this alternative hypothesis.

METHOD

The percentage of items keyed for socially undesirable responses, based upon the social desirability scale values of the MMPI items obtained by Heineman (1953) and reproduced in Dahlstrom and Welsh (1960) was obtained for 31 of the 32 MMPI scales investigated by Couch and Keniston.[2] The product-moment correlation between the percentage of keyed socially undesirable responses in each scale with the correlation of the scale with the OAS was found to be .77. The corresponding product-moment correlation between the percentage of keyed True responses

[1] In view of the definition of a socially desirable response, it should be obvious that a socially undesirable response is defined as a False response to a personality item with a socially desirable scale value or a True response to a personality item with a socially undesirable scale value.

[2] Scoring keys for many of the 32 MMPI scales are given by Dahlstrom and Welsh (1960). Harrison Gough kindly made available to us the scoring keys for the scales not given in Dahlstrom and Welsh. We used only 31 of the 32 scales because for one scale, Social Presence (*X*1), the percentage of keyed True responses reported by Couch and Keniston was not in accord with the percentage given by Gough's key. Two other minor discrepancies were noted in the percentage of items reported as keyed True by Couch and Keniston and the scoring keys given by Dahlstrom and Welsh. Couch and Keniston report that 74% of the items in the *Sc* scale are keyed True and 35% of the items in the *D* scale are keyed True. The percentages, based upon the scoring keys given by Dahlstrom and Welsh are 76 and 33. These two minor discrepancies would in no way seriously influence any of the results reported by Couch and Keniston or in the present paper.

in each scale and the correlation of the scale with the OAS was .85.

For the 31 scales the correlation between the percentage of keyed True items and the percentage of keyed socially undesirable items is .55. The partial correlation between the percentage of socially undesirable items and the OAS correlations, with the percentage of True items held constant, is .69. The corresponding partial correlation between the percentage of True responses and the OAS correlations, with the percentage of socially undesirable items held constant, is .80.

DISCUSSION

On the basis of the results reported in the present study, the OAS correlations with the MMPI scales are consistent with both the hypothesis that scores on the OAS (and MMPI scales) are influenced by acquiescent tendencies and the hypothesis that they are influenced by social desirability tendencies. There is no compelling evidence in favor of either hypothesis.

It is unfortunate that the possible confounding of the acquiescence and social desirability sets was not given consideration in the development of the OAS. An acquiescence scale independent of the tendency to give socially desirable (or socially undesirable) responses would be of obvious value in personality research.

REFERENCES

COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, **60**, 151–174.

COWEN, E. L., & TONGAS, P. The social desirability of trait descriptive terms: Applications to a self-concept inventory. *J. consult. Psychol.*, 1959, **23**, 361–365.

DAHLSTROM, W. G., & WELSH, G. S. *An MMPI handbook*. Minneapolis: Univer. Minnesota Press, 1960.

EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.

EDWARDS, A. L. Social desirability and *Q* sorts. *J. consult. Psychol.*, 1955, **19**, 462.

EDWARDS, A. L. Social desirability and probability of endorsement of items in the interpersonal check list. *J. abnorm. soc. Psychol.*, 1957, **55**, 394–396. (a)

EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957. (b)

EDWARDS, A. L. Social desirability and the description of others. *J. abnorm. soc. Psychol.*, 1959, **59**, 434–436.

EDWARDS, A. L., HEATHERS, LOUISE B., & FORDYCE, W. E. Correlations of new MMPI scales with Edwards' *SD* scale. *J. clin. Psychol.*, 1960, **16**, 26–29.

HANLEY, C. Social desirability and responses to items from three MMPI scales: *D*, *Sc*, and *K*. *J. appl. Psychol.*, 1956, **40**, 324–328.

HEINEMAN, C. E. A forced-choice form of the Taylor anxiety scale. *J. consult. Psychol.*, 1953, **17**, 447–454.

HILLMER, M. L., JR. Social desirability in a two-choice personality scale. Unpublished master's thesis, University of Washington, 1958.

KENNY, D. T. The influence of social desirability on discrepancy measures between real self and ideal self. *J. consult. Psychol.*, 1956, **20**, 315–318.

TAYLOR, J. B. Social desirability and MMPI performance: The individual case. *J. consult. Psychol.*, 1959, **23**, 514–517.

WIGGINS, J. S., & RUMRILL, C. Social desirability in the MMPI and Welsh's factor scales *A* and *R*. *J. consult. Psychol.*, 1959, **23**, 100–106.

(Early publication received November 7, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 175–179

AGREEING RESPONSE SET AND SOCIAL DESIRABILITY

ARTHUR COUCH AND KENNETH KENISTON
Harvard University

TAYLOR (1961) and Edwards and Walker (1961) have suggested that our previous study of "Yeasayers and Naysayers" (Couch & Keniston, 1960) should be re-examined for manifestations of the social desirability variable (Edwards, 1957). This note reports further analyses of our data to this end.

Edwards (1957) has defined the social desirability variable as the tendency of subjects to answer "Yes," "True," or "Agree" to statements having consensually-defined social desirability, and to answer "No," "False," or "Disagree" to statements having consensually-defined social undesirability. Two different operational approaches have been used to determine the degree of social desirability of personality test items: (*a*) a panel of judges rates each item in terms of this dimension, then these ratings are scaled by appropriate methods to arrive at the social desirability value of the statement; (*b*) the probability of endorsement of the item by a population of subjects can serve as an inductive indicator of the same variable: high frequency of endorsement defines a socially desirable item, while low frequency of endorsement defines a socially undesirable item. The high relationship between these two methods of assessment has been amply demonstrated by Edwards (1957) and others.

Edwards attempted to develop a short measure of this general response tendency by having a panel of 10 judges evaluate a set of MMPI items from the *F*, *L*, *K*, and *A* scales (Edwards, 1957, pp. 29–30) for the social desirability of a "True" or "False" answer. The 79 items with perfect agreement as to social desirability were subjected to further item analysis. The subsequent selection resulted in a 39-item scale that Edwards has termed the Social Desirability scale (*SD*).

Relationship between the SD and the OAS

The *SD* (Edwards, 1957, p. 32) score was computed for the 61 subjects (Harvard sophomores) in our previously reported study. The mean *SD* score was 29.6 with a standard deviation of 6.4. These values are quite similar to those reported by Edwards (1957).

In our previous article (Couch & Keniston, 1960, pp. 152–153) we described our procedures for assessing the agreeing response tendency from the sum of Likert-style responses on an Agree to Disagree continuum to 360 personality test items that had been carefully selected for diversity of content. This Over-all Agreement Score (OAS) provided our basic measure of the agreeing response set.

Taylor (1961) and Edwards and Walker (1961) both suggest that our OAS measure may be highly negatively related to social desirability. They hypothesize that subjects high on the OAS ("yeasayers") are endorsing socially *un*desirable items, while those low on the OAS are endorsing socially desirable items (and rejecting the undesirable ones).

In order to test this hypothesis, the correlation between the OAS and the *SD* was computed for our population. A coefficient of $r = -.34$ resulted. While this correlation is significant at the .01 level in the predicted direction, there are three points to be considered in interpreting the finding. First, the level of the correlation with the *SD* is relatively low in comparison with the level obtained with other personality variables that are highly related to the OAS; for example, the Anal Preoccupation scale ($r = .75$), the Excitability scale ($r = .57$), and the Conventional Anxiety scale ($r = .57$) (Couch & Keniston, 1960, p. 157). Second, the *SD* is so constructed that approximately 77% of the items are keyed "False." Thus, a negative correlation with the OAS would be expected on the basis of the agreeing response tendency alone. Third, it is pertinent to question whether the *SD* actually represents a pure and valid measure of the social desirability variable. The *SD* was developed from a rather limited array of items on a single test (MMPI) in an area of content centered around the admission or denial of anxiety and mental disturbance. This component of abnormality (which is also socially undesirable) is clearly evident in the following items of the *SD*:

1. My hands and feet are usually warm enough. (True)
2. I am very seldom troubled by constipation. (True)
3. Most of the time I would rather sit and daydream than to do anything else. (False)
4. My sleep is fitful and disturbed. (False)
5. I have had periods in which I carried on activities without knowing later what I had been doing. (False)
6. I am not afraid to handle money. (True)
7. I frequently notice my hand shakes when I try to do something. (False)
8. I sweat very easily even on cool days. (False)
9. I cannot keep my mind on one thing. (False)
10. I feel anxiety about something or someone almost all the time. (False)

11. I have been afraid of things or people that I knew could not hurt me. (False)
12. I feel hungry almost all the time. (False)
13. It makes me nervous to have to wait. (False)
14. I shrink from facing a crisis or difficulty. (False)
15. I sometimes feel that I am about to go to pieces. (False)

The admission or denial of anxiety symptoms is probably, in part, a reflection of the person's actual degree of disturbance vs. normality. Thus, it is not clear what generality the *SD* has as a measure of response set, nor is it clear to what extent the variable of "Ego-strength vs. Anxiety" is permeating the measure and thereby contaminating the scale.

Development of the High Mean and Low Mean Scales

In order to investigate further the relationship between agreeing response set and s ial desirability, alternative measures of social desirability were developed from our own assessment data. Following Edwards' demonstration that the probability of endorsement of items is highly related to their social desirability, we reasoned that items with high means in our population would reflect sentiments which are the most acceptable (or socially desirable) in this cultural setting. Conversely, items with very low means should tend to reflect a lack of social desirability, and, thus, promote a high frequency of disagreement.

From the total set of 360 items in our OAS measure, we selected the 35 items with the highest means to form a High Mean scale (Mean 5.22; *SD* 0.35), and the 35 items with the lowest means to form a Low Mean scale (Mean 2.41; *SD* 0.44). All items in the High Mean scale had means greater than 5.0; while the Low Mean scale contained items with means less than 3.0. The Likert scale was scored: Strongly Agree (7), Agree (6), Slightly Agree (5), No Answer (4), Slightly Disagree (3), Disagree (2), and Strongly Disagree (1). A sample of the items in these scales[1] is presented in Table 1.

[1] The complete set of 360 items in the OAS measure has been deposited with the American Documentation Institute. We have also included the means and standard deviations of each item, as well as an indication of the 35 items in the High Mean and Low Mean scales. Also, the complete set of centroid and rotated factor loadings for all eight factors has been deposited with the American Documentation Institute. Order Document No. 6694 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 1
SAMPLE ITEMS FROM THE HIGH MEAN AND LOW MEAN SCALES

Sample Items	*M*	*SD*
High Mean Scale		
There are days when one awakes from sleep without a care in the world, full of zest and eagerness for whatever lies ahead of him.	5.9	1.0
Happiness is one of the primary goals of life.	5.7	1.3
A man should be his own harshest judge.	5.6	1.1
The world is teeming with opportunities and promises of success for anyone with sufficient imagination to perceive them.	5.6	1.3
Most people you meet are friendly and obliging, more disposed to aid you than to refuse aid.	5.4	1.1
I like to think things out ahead of time.	5.3	1.2
The vast majority of men are truthful and dependable.	5.3	1.1
I get along well with people.	5.2	1.1
I persist in the face of difficulties.	5.1	1.0
Don't encourage fears and anxieties by dwelling on them.	5.1	1.2
Low Mean Scale		
The future looks black as pitch, with little in view to justify a core of hope or faith.	1.8	1.7
It is misery to be born, pain to live, grief to die.	1.8	1.1
Beneath the polite and smiling surface of man's nature is a bottomless pit of evil.	1.9	1.2
I cannot escape the conviction that fate somehow has it in for me.	1.9	1.0
The prospect is pretty hopeless: it looks as if the nations were heading for their doom—one final, fateful, suicidal, global war.	2.1	1.2
Shower your friends with gifts; that is the way to a full life.	2.1	1.0
If a man is to fulfill his destiny he can expend no more than a fraction of his supply of energy in the service of others.	2.2	0.9
No one is of great emotional significance to me.	2.3	1.4
I couldn't care less what others think of me.	2.4	1.1
The only reason for being "moral" is that you usually get caught if you aren't.	2.5	1.4

An inspection of these items lends subjective support to the thesis that a high probability of endorsement (agreement in this case) indicates a high social desirability of the attitude implied by the item, for example, "I get along well with people." and "I persist in the face of difficulties." On the other hand, a low probability of endorsement (disagreement in this case) seems to indicate a low social desirability of the item, for example, "I couldn't care less what others think of me." and "The only reason for being 'moral' is that you usually get caught if you aren't."

A high score on the High Mean scale should measure the positive side of the social desirability tendency, while a high score on the Low Mean scale should measure the opposite tendency to endorse socially undesirable items. The nature of these two measures of social desirability permits a more direct test of the relationship of this tendency to the agreeing response set. Both scales are susceptible to the agreeing response tendency, but differ as to whether this effect is compounded with a positive or negative measure of social desirability. On either the High Mean or the Low Mean scales, a high score should reflect a similar positive contribution of the agreement tendency; however, the scales should involve opposite contributions from the social desirability tendency. Thus, both scales should have positive correlations to the OAS in so far as agreeing response set permeates the measures.

Reasoning on the basis of the social desirability variable, the High Mean and Low Mean scales should have opposite relationships to each other and to the *SD*—since the High Mean scale is a positive measure of the social desirability tendency while the Low Mean scale is a measure of the reversed tendency (social undesirability). If both response set tendencies are operating to equal degrees, then there should be no relationship between the two Mean scales. The computed intercorrelations between these response set scales are presented in Table 2.

Table 2 shows that both the High Mean and Low Mean scales are significantly related in a positive direction (r = .65 and .42, respectively) with the OAS measure. The fact that the correlation with the OAS is considerably higher for the High Mean scale than for the Low Mean scale appears to contradict Taylor's (1961) contention that agreeing response set would be negatively related to the social desirability tendency. The evidence, in fact, points in the opposite direction.

TABLE 2

INTERCORRELATIONS OF RESPONSE SET MEASURES

Scales	OAS	*SD*	High Mean	Low Mean
OAS	1.00	−.34	.63	.42
SD		1.00	−.16	−.38
High Mean			1.00	−.23
Low Mean	(N = 61)*			1.00

* .05 level of significance (r = .26).

The expected negative correlation between the Low Mean scale and the *SD* is found (r = −.38); however, there is a rather surprising lack of relationship between the High Mean scale and the *SD* (r = −.16). In our population, there is no relationship between the tendency to agree with the most frequently endorsed (agreed with) items and scores on the *SD*.

The expected lack of relationship between the High Mean and Low Mean scales is found (r = −.23), thus, lending some support to our hypothesis of opposing components of response sets. However, the explanation of this finding is not clear from the present set of rather confusing intercorrelations. We therefore turned to a more thorough approach using factor analysis.

Factor Analysis of MMPI Scales and Response Set Measures

Much previous research on the determinants of personality test scores has turned up convincing evidence for the presence of two pervasive response sets—namely, the tendency to agree (or disagree) with items regardless of their content, and the tendency to endorse (or reject) items that appear socially desirable. Since the effects of these response sets are frequently compounded with one another, it is difficult to identify their independent correlates in the personality realm or even to decide which response set explains a particular pattern of findings. Having developed measures that assess these tendencies in different ways, it should be more possible to study their unique characteristics in a domain of personality variables.

Since much previous work as well as the present controversy has centered on the personality correlates of response set in the MMPI, it seemed appropriate to factor analyze a combined domain of MMPI scales and response set measures. For each subject, we had scores for 32 MMPI scales that measured personality variables, a Sum True score that represented the total number of "True" responses given on the MMPI, the *SD*, the OAS, and the High Mean and Low Mean scales. These 37 variables were intercorrelated and the resulting correlational matrix was factor analyzed by the Thurstone Complete Centroid method. Eight factors were extracted which altogether accounted for approximately 78% of the total variance.

Since our interest is focused on the response set

tendencies, the first two factors were rotated through the main measures under discussion: namely, the *SD* and the OAS. A third factor was positioned in reference to the High Mean and Low Mean scales. Subsequent factors have been disregarded in this analysis.[1] In Table 3 we have presented these rotated orthogonal factor loadings.

The very high loadings of the SDS on Factor I (.88) and the OAS on Factor II (.87) allow us to conclude that the first two dimensions clearly reflect the main aspects of these response set measures. It should be noted that the OAS measure is not loaded on the SD factor, and the *SD* measure has only a small loading (−.38) on the OAS factor. Since the factors are orthogonal, this pattern implies that the response set tendencies reflected in these measures are essentially independent of each other.

TABLE 3

ROTATED FACTOR LOADINGS OF MMPI AND RESPONSE SET SCALES

Scales[a]	I	II	III	h^{2b}
1–*L*	.06	−.40	−.17	.61
2–*F*	−.50	.15	−.23	.78
3–*K*	.64	−.46	−.06	.86
4–*Hs*	−.55	.30	.05	.64
5–*D*	−.61	.22	−.41	.86
6–*Hy*	−.25	−.09	.09	.58
7–*Pd*	−.49	.18	−.27	.61
8–*Mf*	−.23	.08	.18	.31
9–*Pa*	−.44	.12	.11	.70
10–*Pt*	−.78	.53	−.09	.94
11–*Sc*	−.80	.40	−.13	.98
12–*Ma*	−.22	.44	.17	.76
13–*St*	.55	−.05	.32	.80
14–*Pr*	−.45	.62	−.34	.97
15–*Do*	.51	−.15	.12	.67
16–*Re*	.32	−.58	.06	.81
17–*Ac*	.63	−.39	.13	.79
18–*Sp*	.76	.07	.06	.82
19–*Ie*	.59	−.26	.12	.84
20–*Ai*	.22	−.47	.13	.75
21–*Si*	−.80	.04	−.15	.88
22–*Es*	.59	−.35	.20	.78
23–*A*	−.76	.54	−.13	.95
24–*R*	−.36	−.44	−.27	.62
25–*Dp*	−.69	.58	−.05	.96
26–*Fm*	−.47	.46	−.29	.64
27–*Rp*	.56	−.44	−.07	.73
28–*To*	.46	−.61	.33	.98
29–*X1*	.58	−.02	.25	.74
30–*X2*	−.45	.55	.06	.73
31–*Oi*	.06	−.27	.19	.50
32–*Lp*	.76	−.36	−.01	.90
33–Sum True	−.43	.74	.06	.99
34–*SD*	.88	−.38	.00	.92
35–OAS	.00	.87	.00	.76
36–High Mean[c]	.16	.77	.55	.92
37–Low Mean	−.33	.23	−.58	.52
Σa^2 =	10.56	6.74	1.82	28.72

[a] The MMPI scales are given standard abbreviations. These are explained more fully in our previous report (Couch & Keniston, 1960, p. 155).

[b] The communalities are for eight factors.

[c] An additional High Mean scale of 50 items has been omitted from this analysis.

The Factor I loadings on the MMPI scales correspond closely to the same pattern of relationships that Edwards (1957) has previously noted as the correlates of the *SD*. In terms of previous factor analytic studies of the MMPI (Kassebaum, Couch, & Slater, 1959), this "SD dimension" seems to represent the general disposition of Anxious Social Withdrawal vs. Normal Social Participation.

The Factor II loadings of the MMPI scales show very much the same pattern as the correlations of these scales with the OAS measure, as reported in our previous article (Couch & Keniston, 1960, p. 155). The general dimension might best be termed Emotional Impulsivity vs. Control.

Of particular interest are the loadings of the several response set measures on the first two factors. The Sum True measure from the MMPI is negatively related (−.43) to the SD Factor I and positively related (.74) to the OAS Factor II. The tendency to answer "True" on the MMPI apparently involves both the tendency to endorse socially undesirable admissions of anxiety and the agreeing response tendency itself.

As already noted, the *SD* has a small negative loading (−.38) on the OAS Factor I. This part of the *SD*'s variance may be that portion due to the high percentage of "False" responses on the scale.

In view of the present controversy, it is quite significant that the High Mean and Low Mean scales do not have any substantial loadings on the first SD factor. The only high loading of these scales on the first two dimensions is the positive one (.77) of the High Mean scale on the OAS Factor II. Apparently, agreement with "socially desirable" items is a relatively good measure of the agreeing response tendency; whereas, agreement with "socially undesirable" items is a far less appropriate indicator of that same tendency.

The lack of substantial loadings of the High Mean and Low Mean scales on the SD Factor I make it clear again that these new response set scales of social desirability are not highly related to the SD measure. This finding throws some doubt upon the validity of the *SD* as a general measure of the social desirability variable.

The most interesting new finding is that the High Mean and Low Mean scales emerge as the clear defining points of a third bipolar factor. The high positive vs. negative loadings suggest that these two scales are measuring some bipolar

response tendency related to the "endorsement" variable used in their method of construction. The best measure of this third factor would be a Balanced (High–Low) Mean scale that would cancel out the agreeing response set. The factor loadings of this scale as estimated by Thomson's Pooling Square technique are:

	I	II	III	h^2
Balanced Mean scale	.33	.34	.72	.74

The loadings of this Balanced Mean scale on the first two factors indicate quite low relationships to the general dimensions reflected in the SDS and the OAS measure. The high loading on the independent third dimension suggests that these new scales may represent a major aspect of the social desirability tendency that is not related to the *SD*.

SUMMARY

Some questions have been raised concerning the relationship of the agreeing response tendency to the social desirability variable. In an attempt to investigate this problem we have presented evidence suggesting that agreeing response set as measured by the OAS is independent of whatever aspect of the social desirability tendency is measured by the Social Desirability scale.

Certain new measures—the High Mean and Low Mean scales—were developed to assess the social desirability tendency, using the degree of endorsement by the subject population as a criterion of construction. These inductive measures of social desirability were not highly related to the *SD*.

A factor analysis of the several response set measures and an array of standard MMPI scales supports our contention that the agreeing response tendency and the social desirability variable are independent of one another. The emergence of a third orthogonal factor defined by the High Mean and Low Mean scales suggests the possibility that a major aspect of the social desirability tendency is unrelated to the *SD*.

Although we have not carefully examined the personality correlates of the different social desirability measures in this paper, the factor analysis implies that the SD measure is centered around the admission or denial of anxiety, while the High Mean and Low Mean measures are centered around a conformity to conventional values.

The manifestation of different response sets in the area of personality tests is quite obviously a challenging and complex problem—and well merits further examination by interested researchers.

REFERENCES

COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, **60**, 154–174.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

EDWARDS, A. L., & WALKER, J. N. A note on the Couch and Keniston measure of agreement response set. *J. abnorm. soc. Psychol.*, 1961, **62**, 173–174.

KASSEBAUM, G. G., COUCH, A. S., & SLATER, P. E. The factorial dimensions of the MMPI. *J. consult. Psychol.*, 1959, **23**, 226–236.

TAYLOR, J. B. The "yeasayer" and social desirability: A comment on the Couch and Keniston paper. *J. abnorm. soc. Psychol.*, 1961, **62**, 172.

(Received January 12, 1961)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 180-183

SOCIAL DESIRABILITY AND AGREEMENT RESPONSE SET

ALLEN L. EDWARDS AND JERALD N. WALKER
University of Washington

In view of the considerable body of research published during the past 9 years on the pervasive influence of the social desirability variable on scores on personality scales of the self-report type, we regard it as unfortunate that our note (Edwards & Walker, 1961) was necessary in order to call the attention of Couch and Keniston (1961) to the possible confounding of social desirability and acquiescence in their OAS.

In their original article, Couch and Keniston (1960) found that MMPI scales with a large percentage of items keyed True tended to correlate positively with the OAS while MMPI scales with a large percentage of items keyed False tended to correlate negatively with the OAS. They interpreted these findings as "... indicating that scales with an unusually high (or low) percentage of items scored True are significantly influenced by the agreeing response set" (p. 154) which they regard the OAS as primarily measuring.[1]

In our note (Edwards & Walker, 1961) we called attention to an alternative hypothesis: that scores on both the OAS and MMPI scales may also be influenced by the tendency of subjects to give socially desirable (or undesirable) responses. Using their data, we reported a correlation of .77 between the OAS correlations with the MMPI scales and the percentage of keyed socially undesirable responses in the MMPI scales. The corresponding correlation between the OAS correlations with the MMPI scales and the percentage of keyed True items in the MMPI scales was .85. On the basis of these results, we concluded that the OAS correlations with the MMPI scales were consistent with both the hypothesis that scores on the OAS (and MMPI scales) are influenced by social desirability tendencies *and* the hypothesis that they are influenced by acquiescent tendencies. We also stated that there was no compelling evidence in favor of either hypothesis.

After reading their second article (Couch & Keniston, 1961), we arrive at exactly the same conclusion and point of view we expressed in our note. We still believe, for reasons developed below, that they have not presented any conclusive evidence in favor of either hypothesis.

[1] If this statement is taken seriously, then one may ask why it is that the OAS correlates only .01, .16, .19, .10, and .14 with five of the Thurstone Temperament scales in which the corresponding proportion of keyed Yes responses is 1.00, .95, .90, .80, and .70, respectively. Couch and Keniston offer no explanation for these "exceptions."

CORRELATIONS OF ARS WITH THE *SD*

OAS Scale

Couch and Keniston find that the OAS scale correlates negatively and significantly with the *SD*. They suggest that the negative correlation can be accounted for by the fact that 77% of the items in the *SD* are keyed False; thus, a negative correlation is to be expected with the OAS on the basis of the agreeing response tendency operating in the *SD*.[2] We suggest, as an alternative hypothesis, that the negative correlation is the result of the fact that subjects who obtain high scores on the OAS are giving, in general, somewhat more socially undesirable responses to the items in the OAS than subjects with low scores. In other words, we would predict that the OAS contains more items with socially undesirable scale values than items with socially desirable scale values. If this is the case, then a subject who has a tendency to give socially desirable responses would be more likely to disagree with the items with socially undesirable scale values and, thus, obtain a low score on the OAS. Do the facts support this alternative hypothesis?

Social desirability scale values are not available for all 360 items in the OAS, but we do have the mean ratings assigned the items on the seven-point agreement scale. The proportion of the total number of items with mean ratings in each scale category is as follows: Strongly Agree (.000), Agree (.000), Slightly Agree (.105), No Answer (.410), Slightly Disagree (.321), Disagree (.152), and Strongly Disagree (.011). Thus, none of the items in the OAS have mean ratings in the Strongly Agree, or Agree category. The distribution of these ratings, in view of the consistently found relationship between probability of endorsement of an item and its social desirability scale value, suggests to us that the OAS is more heavily weighted with with socially undesirable scale values than it is with items with socially desirable scale values. The negative correlation of the OAS with the *SD* is compatible with this analysis.

One may raise the question, as Couch and Keniston do, as to why the negative correlation, although significant, is not higher than it is. It has been suggested elsewhere (Edwards, 1957; Ed-

[2] The problem of acquiescence in the *SD* has not been ignored. It was first discussed by Edwards (1957, pp. 34–39) and evidence was presented bearing upon the issue. A more detailed discussion with further evidence is given by Edwards (1961).

wards, Heathers, & Fordyce, 1960) that relatively low correlations of personality scales with the *SD* may be obtained when the scale has (*a*) a relatively large number of items with neutral social desirability scale values, or (*b*) a fair degree of balance in the social desirability keying of the items. Assuming, as Couch and Keniston do, that the mean rating assigned the items on the agree response continuum is some indication of the social desirability scale value of the item, we note that .836 of the total number of items have mean ratings in the middle three categories (Slightly Agree, No Answer, and Slightly Disagree) of their continuum. No items have mean ratings in the two extreme (socially desirable) Agree and Strongly Agree categories and, thus, the remaining relatively small proportion, .164, of the items fall in the two extreme (socially undesirable) Disagree categories. If the distribution of items had been symmetrical about the middle category, we believe the correlation between the OAS and the *SD* would have been lower than the reported value. It is the loading of the OAS with many neutral items and the slight weighting in favor of the Disagree items that we believe results in the relatively low negative correlation of the OAS with the *SD*.

High Mean Scale

The High Mean scale is found by Couch and Keniston to have a low and insignificant negative correlation with the *SD*. They argue that since the 35 items in this scale all have "high" mean agreement ratings, then they must also have "high" social desirability scale values. Of importance here is what is meant by "high." The mean agreement rating of the 35 items in the High Mean scale is 5.22 and this is their Slightly Agree category. None of the items have mean ratings in the Agree or Strongly Agree category. It seemed questionable to us, therefore, whether one could assume that all of these items had "high" social desirability scale values, and we decided to have the 35 items rated for social desirability.

We presented the 35 items in the High Mean scale to 46 students at the University of Washington and obtained ratings of social desirability for each item on a nine-point scale, using methods described by Edwards (1957). The mean social desirability rating of the 35 items on the nine-point scale was 6.3 with a standard deviation of 1.66. This category on the social desirability continuum is labeled "Mildly Desirable" and is adjacent to the "Neutral" category. We believe, on the basis of these findings, that Couch and Keniston are incorrect in their belief that the items in the High Mean scale all have "high" social desirability scale values and that, therefore, the High Mean scale should have a high positive correlation with the *SD*. As we shall show later, we believe, because of the nature of the items in the High Mean scale, that it is a better measure of the agreement response set than either the OAS or the Low Mean scale.

Low Mean Scale

That a negative and significant correlation should be obtained between the Low Mean scale and the *SD* is again consistent with the fact that the 35 items in this scale have mean agreement ratings in the two *extreme* (Disagree and Strongly Disagree) categories and, thus, this scale may be expected to be more sensitive to social desirability tendencies. We suspect, however, that if the items in the Low Mean scale were scaled for social desirability some of them would be fairly close to the neutral section of the social desirability continuum and that possibly some would fall on the socially desirable side of the continuum. It is possible, in other words, that there is also some degree of balance in the social desirability keying of the items in this scale.

THE FACTOR ANALYSIS

As anyone who has ever worked with factor analytic techniques knows, rotations, unless one is seeking simple structure, are completely arbitrary. A plot of the rotated factor loadings given by Couch and Keniston will show convincingly that their choice of a rotation did not result in simple structure and, for that matter, neither does the rotation we have chosen to make.[3] We leave it to the reader, therefore, to decide whether the rotation made by Couch and Keniston makes better psychological *and* psychometric sense than the rotation we have made. We believe that it would be very informative for the reader to make, in the best Thurstone tradition, the actual plots of the rotated factor loadings as given by us in Table 1 and as given by Couch and Keniston.

The loadings of various reference scales on our two orthogonal factors are given in parentheses in our discussion with the loading on the first factor given first and the second factor second. The positive pole of our first factor is defined by the *SD* (.90, .00) and various other scales which have been shown in the past to have high positive correlations with the *SD*. Similarly, the negative pole is defined by various MMPI scales which

[3] We chose a rotation that would maximize the loading of the *SD* on the first factor and minimize it on the second, whereas Couch and Keniston chose a rotation which would maximize the OAS on the second factor and minimize it on the first.

TABLE 1
ROTATED FACTOR LOADINGS OF MMPI AND RESPONSE SET SCALES

Scales	I	II
L	.13	−.56
F	−.63	−.33
K	.73	−.36
Hs	−.70	−.03
D	−.71	−.41
Hy	−.22	−.37
Pd	−.55	−.19
Mf	−.07	.06
Pa	−.48	−.25
Pt	−.92	.14
Sc	−.94	−.10
Ma	−.40	.47
Si	.65	.37
Pr	−.82	.28
Do	.59	.03
Re	.67	−.34
Ac	.81	.02
Sp	.68	.37
Ie	.79	.13
Ai	.60	−.18
Si	−.69	−.29
Es	.76	.10
A	−.91	.16
R	−.11	−.66
Dp	−.86	.22
Fm	−.66	.01
Rp	.66	−.30
To	.79	−.29
X1	.63	.52
X2	−.61	.44
Oi	.33	−.15
Lp	.85	.00
Sum True	−.74	.55
SD	.90	.00
High Mean (omitted)[a]	.12	.78
OAS	−.45	.62
High Mean	−.17	.76
Low Mean	−.46	−.09
Σa^2	15.96	4.87

[a] This High Mean scale was included in the Couch and Keniston (1961) factor analysis, but omitted from their discussion.

have been shown in the past to have high negative correlations with the *SD*. We believe our first factor can be simply and accurately described as a social desirability factor. Further evidence on this point is given by Fordyce (1956) and Edwards and Heathers (1961).

The negative pole of the second factor is defined by two scales, *R* (−.11, −.66) and *L* (.13, −.56), two scales which in the past have been found to have relatively low correlations with the *SD*.[4] It is of significance to note that the *R* scale contains 40 items and that *all of the items are keyed False* and that the 15 items in the *L* scale *are also all keyed False*. We believe that the second factor can be described as acquiescence and that *R* and *L* define the negative pole of the factor. The positive pole of the second factor is defined by the Couch and Keniston High Mean scale (−.17, .76) and by another High Mean scale (.12, .78) consisting of 50 items which they included in their factor analysis but omitted from their discussion.

Our original note stated that we believed that scores on the OAS might possibly be the result of both acquiescence and social desirability. The rather substantial loadings of the OAS (−.45, .62) on both the social desirability and acquiescence factor support our original contention.

[4] Psychometric explanations of why *R* and *L* should and do correlate to a relatively low degree with the *SD* are presented by Edwards, Heathers, and Fordyce (1960) and in greater detail by Edwards (1961).

CONCLUSION

One may, of course, speculate at considerable length as to whether what has been consistently identified, in considerable research during the past 9 years, as a social desirability variable is *really* Anxious Social Withdrawal vs. Normal Social Participation, Psychoticism vs. Normality, Ego-Strength vs. Ego-Weakness or some other equally dynamic variable. We do not believe that these and various other reinterpretations of the social desirability variable serve to clarify matters in any degree, although the new labels may be more satisfying to the psychologist who is clinically and dynamically oriented.

We have nothing against personality dynamics, but we do not believe they should be invoked when relatively simple psychometric considerations are adequate to account for the correlations that have been found between various personality scales and the *SD*. And, just as the correlations of personality scales with the *SD* can be accounted for in terms of the psychometric properties of the scales (Edwards, 1957, 1961; Edwards, Heathers, & Fordyce, 1960) so also can the first factor loading of the MMPI scales be accounted for in terms of the zero-order correlations of the MMPI scales with the *SD* (Edwards & Heathers, 1961).

REFERENCES

COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, **60**, 151–174.

COUCH, A., & KENISTON, K. Agreeing response set and social desirability. *J. abnorm. soc. Psychol.*, 1961, **62**, 175-179.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

EDWARDS, A. L. Social desirability or acquiescence in the MMPI? A case study with the *SD* scale. *J. abnorm. soc. Psychol.*, 1961, in press.

EDWARDS, A. L., & HEATHERS, LOUISE B. The first factor of the MMPI: Social desirability or ego-strength? *J. consult. Psychol.*, 1961, in press.

EDWARDS, A. L., HEATHERS, LOUISE B., & FORDYCE, W. E. Correlations of new MMPI scales with Edwards *SD* scale. *J. clin. Psychol.*, 1960, **16**, 26–29.

EDWARDS, A. L., & WALKER, J. N. A note on the Couch and Keniston measure of agreement response set. *J. abnorm. soc. Psychol.*, 1961, **62**, 173–174.

FORDYCE, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, **20**, 171–175.

(Early publication received February 7, 1961)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 184-186

A NOTE ON ROKEACH'S THEORY OF PREJUDICE

HARRY C. TRIANDIS

University of Illinois

Rokeach, Smith, and Evans (1960), in Rokeach's stimulating book *The Open and Closed Mind*, state that "*insofar as psychological processes are involved*, belief is more important than ethnic or racial membership as a determinant of social discrimination" (p. 135). To prove their point, they asked subjects to respond to a nine-point scale ranging from "I *can't* see myself being friends with such a person" to "I can *very easily* see myself being friends with such a person." The stimulus persons used were persons of the same or different race or religion as the subjects, who had the same or different beliefs as the subjects. All possible combinations of race and belief were used. The beliefs dealt with God, communism, labor unions, socialized medicine, etc. The writers state that "most of the time they (the subjects) discriminate on the basis of belief and not on the basis of racial or ethnic group when they are given the opportunity to react to social stimuli differing simultaneously on both characteristics" (p. 153). Thus, they suggest that the subjects prefer people of a different race or religion who agree with them on questions of belief over people of the same race or religion who disagree with them.

If Rokeach et al. were limiting their argument to friendship choices there would be no reason to quarrel with their interpretation of their findings. The view that friendship choices are based on belief similarity has been suggested by other writers (e.g., Newcomb, 1953, 1956, 1958). Triandis (1959) has found that similarity in aspects of the belief systems of subordinates and supervisors is related to interpersonal liking between them. But when Rokeach et al. discuss their findings under the title of "prejudice" the present writer must object. A concept more general than prejudice is the concept of social distance. Rokeach's research technique used friendship as the variable under investigation, and so inevitably dealt only with a relationship involving small social distance. Prejudice and discrimination are much more relevant to acceptance or rejection of a relationship involving relatively large social distance, such as acceptance of a person as a neighbor or as a student in one's university. In the present study, Rokeach's hypothesis was tested by means of a questionnaire covering a wide range of social distances and was not supported by the evidence.

Method

A questionnaire was used that involved a 100-point scale of social distance ranging from willingness to marry a person (0 social distance) to willingness to participate in the lynching of the person (social distance of 97 points). This scale was developed by the present writer in a previous study (Triandis & Triandis, 1960). The scale values were computed using the Thurstone successive intervals procedure, which yields an equal interval scale. The scale included statements concerning the acceptance of the person as an intimate friend (social distance of 11 points), as a roommate (30 points), as a personal chum in the respondent's club (31 points), as a speaking acquaintance (52 points), and also statements referring to the unwillingness of the respondent to accept the stimulus person as a neighbor (70 points), or "in our universities" (81 points).

The subjects were presented with descriptions of 16 stimulus persons arranged in a factorial design. The 16 stimuli appear in Table 1. Each stimulus included the specification of the race, religion, belief system, and occupation of the person who was to be judged on the scale of social distance. Two values of race (Negro and white), two values of religion (same denomination as the subject and different religion from the subject[1]), two values of occupation (bank manager and coal miner), and two values of belief system were used. The two values of belief system were obtained by asking the subjects to read the 13 ways of life that were prepared by Morris (1956). The subjects were told that they would be examined on their content later and were asked to indicate the most preferred and the most rejected way of life. The subjects were then instructed to think of the most preferred way of life when the words "has the same philosophy as you" appeared in the questionnaire and to think of the most rejected way of life when the words "has a different philosophy" appeared.

For each of the 16 stimulus persons, subjects responded to the 15 items of the social distance questionnaire. These responses were scored as in Triandis and Triandis (1960, p. 112). Thus, for each stimulus person a single score representing the social distance experienced towards "him" was obtained. The 16 scores obtained from each of the 112 subjects were subjected to an analysis of variance.

The subjects were 112 University of Illinois white undergraduates of both sexes.

Results

The means and standard deviations obtained from the 112 subjects to the 16 stimuli appear in

[1] Roman Catholic, Protestant, or Jewish denominations, depending upon which appeared to the subject to be the most different from his own religion.

Table 1. The analysis of variance of these mean scores is shown in Column 1 of Table 2 and that of the 1,792 scores obtained from the 112 subjects is shown in Table 3.

Table 1 shows, as expected, that the same philosophy, same religion, white bank manager is acceptable as an intimate friend while the different religion, different philosophy, Negro coal miner is, on the average, not accepted as any kind of friend, although he is accepted in the same apartment house. The large standard deviations of the Negro stimuli reflect the exclusion of all Negroes from the neighborhood (scale value 70) in the case of many subjects, while the majority of subjects accepted some Negroes as friends of some sort (scale value 41 or lower). It is interesting that the Negro coal miner of the same philosophy and religion has the highest standard deviation, a fact that indicates considerable ambivalence among subjects.

TABLE 1

MEAN SOCIAL DISTANCE SCORES OBTAINED FOR EACH STIMULUS

Stimulus Person	Mean of 112 Subjects	*SD*
White, same philosophy, same religion, bank manager	11.1	0.014
Same philosophy, white, bank manager, different religion	12.5	0.39
Same philosophy, white, coal miner, same religion	17.0	2.22
Different philosophy, white, bank manager, same religion	16.2	0.67
Same philosophy, white, coal miner, different religion	18.7	2.41
Different philosophy, white, bank manager, different religion	20.0	1.46
Same philosophy, Negro, bank manager, same religion	21.6	3.98
Different philosophy, white, coal miner, same religion	23.5	3.28
Same philosophy, Negro, bank manager, different religion	27.2	5.36
Different philosophy, white, coal miner, different religion	29.6	4.00
Different philosophy, Negro, bank manager, same religion	31.5	4.04
Same philosophy, Negro, coal miner, same religion	32.3	6.00
Different philosophy, Negro, bank manager, different religion	35.6	4.95
Same philosophy, Negro, coal miner, different religion	37.1	5.92
Different philosophy, Negro, coal miner, same religion	39.0	4.83
Different religion, Negro, coal miner, different philosophy	45.8	4.96

TABLE 2

ANALYSIS OF VARIANCE OF MEAN SCORES

Source	All Subjects		Other Subjects		Open-minded Subjects	
	SS	%	*SS*	%	*SS*	%
Race	2.71	59	3.68	64	1.93	50
Occupation	.83	18	.82	14	.85	22
Philosophy	.74	16	.79	14	.70	18
Religion	.22	5	.22	4	.22	6
Interactions	.12	2	.23	4	.17	4
Total	4.62		5.74		3.87	

TABLE 3

ANALYSIS OF VARIANCE OF INDIVIDUAL SCORES

Source	*SS*	%	*df*	Variance	*F*	*p*
Race	302.70	15.7	1	302.70	378	<.001
Philosophy	83.27	4.3	1	83.27	104	<.001
Occupation	93.21	4.9	1	93.21	115	<.001
Religion	24.26	1.2	1	24.26	30	<.001
Subjects	1415.60	73.5	1776	.80		
Interactions	7.33	.3	11	—		
Total	1926.37		1791			

Table 3 shows that while all treatments produce significant differences in social distance, race is by far the most important determinant, in the sense that it controls more variance than the other factors. Race is about four times as important as belief system, which is about as important as occupation. The analysis of variance that is based on the means of the 112 subjects for the 16 stimulus persons (Table 2) shows similar results.

Table 1 throws more light on the matter under investigation. We find that neither the person who has all positive characteristics except belief system nor the person who has all positive characteristics except race are acceptable as intimate friends. However, the latter receives a significantly larger ($p < .0001$) mean social distance score than the person who differs in belief system. For all similar comparisons using the data in Table 1 we obtain the same results. There is no doubt that race is a more important determinant of social distance than is belief system.

Many writers in the field of prejudice have argued that people who are characterized by high tolerance for ambiguity and open-mindedness are less prejudiced. Since one of Morris' (1956) ways seems to deal with this matter (Way No. 7: "We should at various times and in various ways accept something from all other paths of life. . . . "), we compared the scores of the 53 subjects who had chosen this way as their most preferred way with those of the 59 subjects who picked other ways. The analyses of variance of these two groups are shown in Columns 2 and 3 of Table 2. The open-minded subjects tend to be

less prejudiced ($p < .20$), yet race is still a very important determinant of the variance of their social distance scores.

It is also of interest to consider whether the results obtained by Rokeach can be replicated with our data. One of Rokeach's approaches was to list the number of times a person with a different race or religion was preferred over a person with a different belief system. To approximate this measure, we compared the social distance due to race and that due to differences in beliefs and tabulated how frequently one was greater than, smaller than, or equal to the other. We found that 41 subjects showed greater social distance due to race than due to belief. The social distances due to race and belief were equal for 37 subjects, and a smaller social distance due to race than due to belief was obtained in 34 cases. The trend goes against Rokeach's findings, but the chi square is not significant. Thus, our data suggest that race is as frequently a factor in social distance as differences in the belief systems of the two persons. But while belief system frequently affects social distance, it does not often involve a *great deal* of social distance. People do not exclude other people from their neighborhood, for instance, because the other people have different belief systems, but they do exclude them because they are Negroes. Rokeach's magnificent attempt at subsuming all the data under one simple notion succeeded only because of the design of his study. Had he considered something more than friendship he would not have obtained his results. It seems that an adequate theory of prejudice is much more complicated than Rokeach's. The author (Triandis & Triandis, 1960) has suggested a theory that can subsume Rokeach's as a special case.

The present study is also open to criticisms. For instance, is the distance between the most preferred and the most rejected value system, in Morris' (1956) scheme, as great a distance between belief systems as one could obtain in life? Probably not. On the other hand, the distance between communist and anticommunist (beliefs used by Rokeach) may in fact be smaller than the distance between some of Morris' ways and yet might have controlled much more variance than the differences in philosophy used in the present study. "Communism" in the present climate of international relations includes "surplus meaning" involving visions of saboteurs, murderers, etc. It seems likely that the use of beliefs involving communism, God, etc. accentuated Rokeach's results and favored his hypothesis. His data dealing with labor unions (p. 141, Table 7.1) were mush less compelling than his data on the other belief systems.

The present note has been quite negative to Rokeach's theory of prejudice. This in no way detracts from the great value of the other parts of his most stimulating book.

SUMMARY

Rokeach's hypothesis that prejudice can be accounted for *primarily* by differences in the belief systems between two persons, or between a person and members of a racial or religious group, was tested and rejected. Rokeach's own experimental confirmation of his hypothesis is due to an artifact of his method of measuring prejudice. His position is re-evaluated in the light of the present findings.

REFERENCES

MORRIS, C. *Varieties of human value.* Chicago: Univer. Chicago Press, 1956.

NEWCOMB, T. M. Motivation in social behavior. In J. S. Brown et al., *Current theory and research in motivation.* Lincoln: Univer. Nebraska Press, 1953. Pp.139–159.

NEWCOMB, T. M. The prediction of interpersonal attraction. *Amer. Psychologist,* 1956, **11,** 575–586.

NEWCOMB, T. M. The cognition of persons as cognizers. In R. Tagiuri and L. Petrullo (Eds.), *Person perception and personal behavior.* Stanford: Stanford Univer. Press, 1958. Pp. 179–190.

ROKEACH, M., SMITH, PATRICIA W., & EVANS, R. I. Two kinds of prejudice or one? In M. Rokeach, *The open and closed mind.* New York: Basic Books, 1960. Pp. 132–168.

TRIANDIS, H. C. Cognitive similarity and interpersonal communication in industry. *J. appl. Psychol.,* 1959, **43,** 321–326.

TRIANDIS, H. C., & TRIANDIS, LEIGH M. Race, social class, religion, and nationality as determinants of social distance. *J. abnorm. soc. Psychol.,* 1960, **61,** 110–118.

(Received December 18, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 1, 187-188

BELIEF VERSUS RACE AS DETERMINANTS OF SOCIAL DISTANCE: COMMENT ON TRIANDIS' PAPER

MILTON ROKEACH
Michigan State University

TRIANDIS' (1961) method is, in my judgment, inappropriate and so also is his conclusion.

The essence of the difference between his method and ours is that while we asked subjects to respond to specific social stimuli such as "Negro who believes in God" and "White atheist," Triandis asks his subjects to respond to "Negro, same philosophy" and "White, different philosophy." Triandis reports that race is now a far more important determinant of social distance than belief, quite the opposite to our findings.

Triandis' conclusion stems not so much from his use of a social distance scale, which he makes so much of, but from his use of the terms "same philosophy–different philosophy," which he makes so little of. What do these terms mean to Triandis' subjects? Triandis tells us that his subjects first studied Morris' (1956) Thirteen Ways of Life, then picked the "most preferred" and the "most rejected" of these 13, and then *imagined* the "most preferred Way" when the stimulus "same philosophy" appeared, and *imagined* the "most rejected Way" when the stimulus "different philosophy" appeared.

By way of example, Morris reports that Way of Life 7 ("preserve the best that man has attained") is generally the most preferred, and Way of Life 13 ("obey the cosmic purposes") the least preferred. These Ways are presented to subjects in the form of paragraphs containing many complex ideas. Each paragraph is over 100 words long. Thus Triandis' subjects must somehow have substituted for "Negro, same philosophy" and "White, different philosophy": "Negro, that-long-and-involved-paragraph-on-Way-7" and "White, that-long-and-involved-paragraph-on-Way-13."

Triandis does not tell us what his subjects remembered of these two paragraphs (not to speak of the 13). Were these paragraphs ego involving? I doubt it. The belief variable pitted against race is vague, abstract, overintellectual, and complicated; hence weak! And the polarities, "same" and "different philosophy," are not really opposites, just different. They are not Ways producing social conflict in everyday life. They represent ideals by which "people say they would like to live, and not how they do live" (Morris & Jones, 1955, p. 534). Thus the difference between "same" and "different philosophy" is a difference which, psychologically and sociologically, makes little difference.

Compare the impact of "Negro, same philosophy" and "White, different philosophy" with the impact of "Negro who believes in God" and "White atheist." Here both race and belief are salient. Both variables are specific, concrete, easy to imagine, and easy to remember. Both variables are ego involving, and it is easy to imagine that such people exist even if one has not personally met them.

Triandis suggests that our results are artifacts due to our selection of a particular socially polarized belief (communism) to pit against the race variable. To make sure our results were not artifacts we repeated our research with eight widely differing issues: God, communism, desegregation, etc. These issues are socially salient and controversial, though not equally so. On all eight issues we found friendship ratings to be more a function of belief than race, though not equally so. These results hold in the South as well as in the North. To these eight issues Triandis now adds a ninth: "same philosophy–different philosophy," and his results are now exactly the opposite. I have tried to state why, in my opinion, this ninth replication is not a crucial one, but rather a trivial one. All our data *and* Triandis' are consistent with the following proposition: the more salient a belief, the more will belief congruence override racial or ethnic congruence as a determinant of social distance.

The preceding considerations notwithstanding, the issue of race versus belief as determinants of discrimination is far from settled. A crucial test of the belief congruence hypothesis will not be forthcoming from more questionnaire studies using different wordings, but from research on what people do in real life. A preliminary study along this line is reported in *The Open and Closed Mind* (1960, p. 156).

Perhaps our findings relevant to the belief congruence hypothesis can be made more palatable if certain ideas previously discussed (p. 164) are rephrased. Only when social constraints are placed on a person would we expect him to discriminate in a way he otherwise would not, if left alone by society. No society constrains *all* the actions of its members, and in those actions not subject to social sanction discrimination along racial or ethnic lines would not take place, not even in the South. In other words, the *locus* of racial and ethnic discrimination is to be sought

in society, not in the individual's psyche. If society's constraints were altogether removed from a particular segment of life man would still discriminate, if discriminate he must, not in terms of race or ethnic grouping, but in accord with his basic psychological predisposition, characteristic of all human beings, to organize the world of people in terms of the principle of belief congruence.

REFERENCES

MORRIS, C. *Varieties of human value.* Chicago: Univer. Chicago Press, 1956.

MORRIS, C., & JONES, L. V. Value scales and dimensions. *J. abnorm. soc. Psychol.*, 1955, **51**, 523–535.

ROKEACH, M. *The open and closed mind.* New York: Basic Books, 1960.

TRIANDIS, H. C. A note on Rokeach's theory of prejudice. *J. abnorm. soc. Psychol.*, 1961, **62**, 184–186.

(Received December 18, 1960)

ERRATUM

On page 145 of the article "Identifying Jewish Names," by P. F. Secord and Ellen Saumer (*J. abnorm. soc. Psychol.*, **61**, 144–145), a mean response bias score of 5.17 for high and low prejudice army personnel is erroneously reported as not significant at the .05 level. This mean is significant at the .01 level, a result which is consistent with previous studies of response bias in the identification of Jews from photographs.